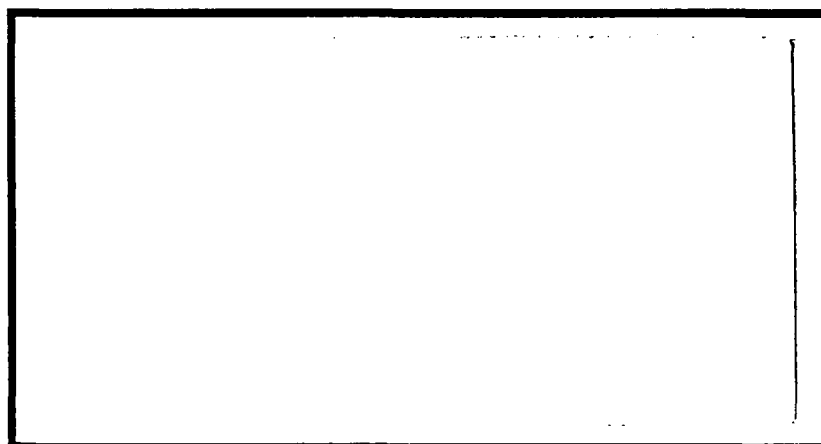


DTIC FILE COPY

1

AD-A202 709



DTIC
ELECTRONIC
JAN 18 1989
S
AD

DISTRIBUTION STATEMENT A
Approved for public release
Distribution Unlimited

DEPARTMENT OF THE AIR FORCE
AIR UNIVERSITY

AIR FORCE INSTITUTE OF TECHNOLOGY

Wright-Patterson Air Force Base, Ohio

89 1 17 039

1

AFIT/GOR/ENS/88D-6

DTIC
ELECTE
JAN 18 1989
S D
DCS

MODELING TELEPHONE ACCESS TO
WILFORD HALL MEDICAL CENTER AND ITS
BUSIEST APPOINTMENT SITES

Martin E. Ellingsworth
Captain, USAF

AFIT/GOR/ENS/88D-6

Approved for public release; distribution unlimited

AFIT/GOR/ENS/88D-6

MODELING TELEPHONE ACCESS TO WILFORD HALL MEDICAL
CENTER AND ITS BUSIEST APPOINTMENT SITES

THESIS

Presented to the Faculty of the School of Engineering
of the Air Force Institute of Technology

Air University

In Partial Fulfillment of the

Requirements for the Degree of

Master of Science in Operations Research

Martin E. Ellingsworth

Captain, USAF

December 1988



Accession For	
NTIS GRA&I	<input checked="checked" type="checkbox"/>
DTIC TAB	<input type="checkbox"/>
Unannounced	<input type="checkbox"/>
Justification	
By	
Date	
Availability Codes	
Dist	Availability Codes
A-1	

Approved for public release; distribution unlimited

Preface

The purpose of this research was to enlighten managers and staff at Wilford Hall Medical Center (WHMC) of the relationships involved in telephone traffic, call queuing and service, and telephone access to health care at the medical center. The need for this research came from the medical center's commander, Major General (Dr.) Vernon Chong, who voluntarily instigated the Ambulatory Care Initiative in WHMC to improve access to health care in line with his motto for the center, "People who care."

The major parts of the telephone system were simulated at the busiest hour of the month. Alternatives for decreasing/eliminating system blocking and increasing the number of calls answered at the decentralized appointment sites were explored. Though the simulation model gave poor fit to the traditional queuing measure of time on-hold due to the high variability in the input data and subsequent model sensitivity, the output variables for system blocking and calls answered gave realistic estimates of performance. The model is ready for further application as more data are collected for better parameterization.

In performing this research and writing this thesis I am indebted to many people. Mr Ron Nagel of the WHMC communications branch and his assistant, Airman Lynn Brown, have been invaluable in my gaining an understanding of the

WHMC telephone system and collecting the data for this study. Colonel Jim Jacobson and Chief Master Sergeant York, of the Ambulatory Care Committee, who have spent many hours revealing how current efforts at WHMC were trying to affect accessibility to care, are both due mention for the help they gave me. A word of thanks is also due to Mr. Bob MacNaughton of WHMC Public Affairs, who helped get me started on this topic and who has been a careful reader of this paper. Last, but not least, of the WHMC support crew, I thank Captain Carl McCann of WHMC/SGMB who saw to it that TDY funds were there to support this work and who has been a constant source of energy in my pursuit of this thesis.

On the AFIT side of the house, Major Joe Litko deserves an award for reading my first draft, and he also can be applauded for his guidance, troubleshooting, and patience during this thesis process. My reader Captain Wade Shaw has also provided much assistance. Finally, I would like to thank my newly wed bride, [REDACTED] for her long-distance support of my work while waiting for my graduation to station us together.

Martin E. Ellingsworth

Table of Contents

	Page
Preface	ii
List of Figures	vi
List of Tables	vii
Abstract	viii
I. Introduction	1
Research Objectives	3
Approach to the Problem	5
Scope and Limitations	8
II. Extended Background	11
The Military Medical Care Environment	12
Wilford Hall Medical Center	14
System Description of WHMC's Phone System	15
The Process Being Modeled	22
The Ambulatory Care Initiative: WHMC and Access to Care	30
Changes in the System Over the Course of This Study	33
Management Update	35
Hypotheses for this Research from the Background Investigation	37
III. Literature Review	42
Telephone Systems and Traffic	42
Determining Teletraffic Demand	44
Telephone System Sizing	47
Telephone System Servicing	49
Modeling Telephone Systems	50
Implementation Issues for Phone Systems	54
Standards of Service in Telephone Systems	55
Discussion of the Literature	60
Findings of the Literature Review and Discussion	68

IV.	Data Reduction	70
	Chapter Synopsis	70
	Definition of Terms	71
	ACD Data	74
	Demand Patterns	85
V.	Methodology	102
	Overview	102
	Modeling	103
	Verification and Sensitivity	109
	Validation	113
	Performance Measurement	118
	Experiments Conducted	119
VI.	Experimentation and Analysis	122
	Overview	122
	Running the Experiments	123
	Changes to the Hardware	123
	Manpower	130
	Service Policies	133
	Scheduling	136
	Further Research Areas	139
VII.	Conclusions and Recommendations	142
	Recapitulation	143
	Findings	144
	Alternatives	146
	Conclusion	157
	Bibliography	159
	Appendix A: A Sample ACD Hourly Report	165
	Appendix B: SLAM Code with Fortran Subroutines	166
	Vita	177

List of Figures

Figure	Page
1. Full System Data	16
2. Telephone System Diagram	18
3. Weekly Call Attempt Pattern	84
4. Full System Data	84
5. Long-term Demand for TOTAL ACD	88
6. Long-term Demand for PCARE	88
7. Long-term Demand for GSURG	89
8. Long-term Demand for GYN	89
9. Long-term Demand for DERM	90
10. Demand Over Time for TOTAL ACD	90
11. Monthly Pattern TOTAL ACD	96
12. Daily Pattern PCARE	96
13. Daily Pattern ORTHP	97
14. Daily Pattern GYN	97
15. Daily Pattern PEDS	98
16. Monday ACD Traffic	98
17. Downtown Minus ACD Traffic	99

List of Tables

Table	Page
I. Input Parameter Calculation	107
II. Baseline Input Parameters for the Model .	108
III. Effect of Time to Renege on Calls Answered and System Blocking	116
IV. Effect of the Number of Commercial Lines on Calls Answered and System Overflow	125
V. Effect of Queue Size on Call Answered and System Overflow	126
VI. Effect of Adding Servers to the Busiest Site	132
VII. Effect of Speed of Service on Calls Answered and System Overflow . . .	135
VIII. Effect of Schedule Changes on Calls Answered and System Overflow	137
IX. Value of Action Matrix	154

Abstract

Telephone traffic into Wilford Hall Medical Center and its busiest appointment centers was simulated at the busiest hour of the month to explore alternatives for improving telephone access to health care. Telephone access to care was defined in two parts: 1) decreasing calls blocked (overflow) of the central commercial switch; and, 2) increasing the number of calls answered at the decentralized appointment sites, which operate with automatic call distributors. Methods for improving current call data reporting procedures were developed. Avenues for feeding back operator performance data to the appointment sites were constructed which may serve as a management control system for monitoring continuous and sustained improvement in the telephone access at each appointment site. The simulation gave insight into how to manage call traffic to decrease the number of commercial lines leased while ensuring a minimum of overflow traffic. Simple changes to the length of the queue of the call distributors and to when new appointments are made available could dramatically decrease the load on the network at the busiest hour of the month. Alternatives for increasing the number of calls answered included dedicated scheduling of operators, automating the appointment making process, and augmented staffing of

operators between the decentralized sites. The strategic problem behind telephone access is that getting to talk to an appointment clerk is not equivalent to getting the appointment for which the call originated. A key question for management is how to deal with calls after all available appointments are distributed for a given clinic. Findings for managing demand traffic and operator performance to improve telephone access to care within the constraints of a not-for-profit health care organization may be useful to other such organizations.

MODELING TELEPHONE ACCESS
TO WILFORD HALL MEDICAL CENTER
AND ITS BUSIEST APPOINTMENT SITES

I. Introduction

Top management within the Wilford Hall Medical Center (WHMC) cares about how people perceive the quality of military medicine. They believe that public perception of the quality of military medical care is influenced by the ease of access to medical care. There is a much greater demand for no-cost medical benefits than the military medical community can meet on a day-to-day basis. The routine demand manifests itself in the form of telephone calls to the various clinic appointment centers throughout the hospital. A key problem exists in how to manage telephone hardware and telephone operator manpower to best serve the patient population in accessing the different appointment centers by telephone. Usable definitions and measures of accessibility, effectiveness, and productivity are needed. Creating these measures, and constructing alternative policies for consolidating or decentralizing appointment centers based on these measures and patient need is the primary thrust of this study.

The telephone system of Wilford Hall Medical Center (WHMC) is a major point of entrance for beneficiaries seeking care at the facility, as well as a primary means of daily communication for patients, visitors and staff alike. It is desirable to be able to conduct routine phone calls in and out of the facility without inconvenience, and to allow calls for appointments to get through to their destinations and be answered without a prolonged wait or numerous busy signals. However, on a typical Monday morning of the first full work week of the month, thousands of calls get blocked from entering the medical center. In addition, most appointment centers experience daily congestion in the morning hours with extreme congestion at the time they book appointments. An excess demand situation exists, and many calls arrive only to find they must call again at the next booking time. This current system does fulfill the task of distributing all the available appointments, but there is still the problem of gaining access to the appointment centers during these peaks in demand.

Those appointment centers with the highest demand are of interest when studying telephone access to the medical center. The appointment site is the first step in accessing the medical center, and calls to these sites account for a large part of the entire telephone demand to WHMC. The busiest appointment sites have special telephone equipment called Automatic Call Distributors (ACD) which permit

keeping many calls on-hold at a time (currently set to six). The ACD software also collects statistics on caller behavior and service by site. The ACD hardware includes a headset which allows the appointment clerks free use of their hands, while the ACD software efficiently forces calls to be answered by automatically connecting a waiting caller with an available clerk at a particular ACD site. Unfortunately, WHMC does not have enough manpower to handle all the calls going to the various ACD appointment sites. Many callers receive a busy signal or an "all circuits busy" message, while callers getting through often experience long waits before either hanging-up, or finally, getting to talk to a clerk. In using a decentralized ACD system with only one or two servers available at any site, WHMC has implicitly defined the level of service it can afford at the present time.

Research Objectives

The objective of this study was to obtain a thorough understanding of the access behavior of the WHMC telephone system with particular attention being paid to the busiest telephone appointment centers. Although quantitative aspects of access are the main concern in this study, qualitative issues are mentioned for the sake of defining possible impacts of selecting a policy for implementation.

It was expected that by studying the system in depth, some useful alternatives might be presented from which to make resource and administrative decisions about telephone hardware, manpower, personnel, training, and service policy. The decisions these alternatives might prompt should enhance the real and perceived telephone accessibility of WHMC and its appointment centers, and should help decrease volume and frequency of caller complaints about telephone access.

The goal of this research was to give WHMC managers insight for use in making decisions related to ACD telephone equipment and staffing. However, not all phone traffic to and from WHMC is governed by the ACD architecture. A simulation model was constructed which accounted for the telephone traffic to the ten ACD sites and also used estimates of the non-ACD traffic in order to mimic the telephone system at WHMC. The effort reduced to looking at the queuing and service characteristics of a large phone system with an embedded, decentralized ACD system. Access was defined as decreasing calls blocked (overflow) to the system while increasing the number of calls answered at the ACD sites. The analysis sought to find how the system behaves, and to search for ways to improve accessibility and limit caller strife.

Subsidiary Objectives. In order to reach the above research objective, the following outline of subsidiary objectives was set forth:

- 1) Assess telephone access to appointment scheduling systems for sites operating on ACD software and hardware.
 - a) Develop demand traffic distributions for each ACD site and describe any peak loading conditions.
 - b) Determine each ACD site's current:
 - 1) operator scheduling and answering policies (e.g. augment staff, when open);
 - 2) appointment log policies; and,
 - 3) ACD hardware and manpower present.
 - c) Determine the demand traffic distribution of non-ACD telephone traffic for its impact on the stacking and blocking of the entire incoming call capability of the WHMC phone system.
- 2) Develop a model to simulate telephone traffic at WHMC, verify and validate it with collected data.
 - a) Create a model containing the processes of call arrival, talk time, and post call work as the data permit for each ACD site and for the non-ACD traffic in WHMC. The model should reflect the individual ACD site's practices, and relate to the total load the incoming lines of the phone system can withstand.
 - b) Evaluate the sensitivity of the model and compare control run results with analytic queuing models to verify the code.
 - c) Compare simulation results with actual data, and see how results compare with other observations and knowledge of the system.
 - d) If the model can be verified and validated, conduct experiments on tradeoffs between caller waiting, the queue size, the number of servers, service time, the number of lines, and the service policy for scheduling and staffing.
- 3) Develop tables, graphs, and/or heuristics for the WHMC system that will aid in the management of telephone resources and service policy. Especially, estimate the number of trunks lines required to serve the system.
- 4) Develop guidelines for collecting and interpreting data, and for updating the parameters in the model, or using them in tables, for management decisions on service policy.
- 5) Aid in developing measures of effectiveness commensurate with WHMC's resources and concerns, and help to set aspiration levels within these measures.

Approach to the Problem

The study began with extensive interviews of managers at various organizational levels in WHMC and with the clerks at each of the major appointment sites. On-site interviews were conducted at increasing depth over the period of study during visits to WHMC in January, March, June, July, and September of 1988. Information from these interviews constructed the framework within which this analysis took place. Two factors were of relevance. The first factor involved the implicit constraints on service resources imposed by the not-for-profit nature of the facility. This presented a problem where manpower was scarce and where there was no cost/revenue structure by which to evaluate any level of service to callers or any alternative system designs. Thus, looking for ways to improve the system as it existed, helping to define a usable service standard, and looking for easy to implement changes to the system or its procedures was the main pursuit of this study.

The second factor was a coinciding initiative in the Air Force medical service known as the Ambulatory Care Initiative (ACI) which started to take hold at WHMC a few months before this study's inception in October 1987. Though broad in scope, part of the direction in the ACI program was to enhance patient access to care. The impact

of this program was the motivation for WHMC sponsoring this study, and the personnel working the ACI program have been active throughout the course of this study implementing many changes and influencing telephone access to the major appointment sites at WHMC. Findings and insights in the access problem were discussed and considered with the study sponsors as they occurred, and the final conclusions will have the attention of a possible implementing agency. While it was impossible to know whether or how this study has directly influenced any actions of the ACI program personnel or the staff in the communications office, who directly supervised this study, it was verifiable that real changes were made to increase the telephone access to care at WHMC during the course of this study. Furthermore, the simulation model and findings from this study may be useful to other Air Force medical centers (14:149), or any large organization with decentralized ACD phone centers, in improving telephone access.

A complete study of the available data on telephone service at WHMC led to several important discoveries on demand patterns for each appointment site. These patterns helped to define variations in telephone traffic and were useful for creating policies for scheduling appointments and staffing operators, and for designing hardware to better handle excess demand. Analytic methods for defining the system were inappropriate due to the system's toleration of

losses and delays, and evident retrials. A simulation model of the existing system during the busiest hour of the month was created to help understand the problem and to evaluate alternative operating conditions.

The system was easy to model, but reducing the available ACD data to useful parameters was difficult. The process was cumbersome, since data was only available in printout format from the ACD system software. Further, the data for similar time periods observed had a great deal of variability making the point parameters for the model a sloppy representation. However, the modeling process was a valuable source of insight, and the final model was useful for comparing operating policies. The study concluded with many people, who prior to the study rarely interacted, having a much more informed and common point of view on the issue of telephone access to WHMC. Ways to improve service within the present organizational structure (or with small changes) were uncovered and implemented, while an opportunity for using further findings was opened at WHMC and possibly at other not-for-profit organizations operating a multi-site, ACD enhanced, telephone access system.

Scope and Limitations

The effort was to construct a simulation model to use in the study of the complex queuing behavior associated with

incoming phone calls to large, multi-department organizations, which have part of their calls serviced by departments on an ACD system. The application of the simulation model was limited to the busiest hour of the month for the telephone system at Wilford Hall Medical Center (WHMC) in San Antonio, Texas. Improving the quantity and speed of service to incoming telephone calls to WHMC while considering clerk efficiency and productivity was the baseline scope of this effort. Various policies for manning the phones and even for consolidating telephone service were investigated. Management, training, and implementation issues were also addressed. However, this study is not a cure-all for the hospital. This study specifically does not address the problem of excess demand for no-cost military medical care, which surfaces as a key issue in patient dissatisfaction -- a caller finally gets through and cannot get the appointment for which they waited. Instead, this study attempted to maximize the quality of service to those who can get in, while minimizing the inconvenience to those callers who have no opportunity to get an appointment in a booking period. This study does not delve into the issues of managing public opinion of military medicine.

The remainder of this paper will associate the reader with the problem in detail and describe some steps for improving the telephone access to WHMC. Chapter Two is an extended background section which introduces the environment

observed during field interviews at Wilford Hall Medical Center over the course of this study. This section also describes the physical layout of the telephone system and gives an overview of observed personnel and organizational issues surrounding the several decentralized telephone appointment centers. Chapter Three contains a review of the literature pertaining to the performance of telephone systems and associated queuing networks. Chapter Four details the specifics of collecting and reducing the data for analysis, and also presents observed telephone traffic patterns for both the routine and the appointment center callers. Chapter Five discusses the methodological approach of the study while listing the assumptions under which the method was chosen. Chapter Six discusses the results of the analyses of the simulation model along with suggestions for further research. Finally, Chapter Seven lists the conclusions and recommendations derived from this study.

II. Extended Background

This chapter gives a description of the military medical health care environment, an overview of WHMC and its phone system, and it portrays the routine situation of the ACD appointment centers. It provides background information that should lead the reader into the complexity and uniqueness of the issues involved at Wilford Hall Medical Center in regard to telephone access. The changes to the system during the timeline over which the study was conducted (October 1987 to November 1988) are introduced. Many of the issues presented here are important for considering the full impact of making changes in the system, and for improving the system. A detailed description of the process being modeled is presented along with some observations made during the interview process. Finally, a service goal is established as maximizing the number of calls answered at the ACD sites while minimizing system overflow. A set of hypotheses about what is needed to help improve service and efficiency at WHMC are formulated from information in the background investigation. Readers not familiar with the environment at WHMC in the last year, should find the details presented next valuable in evaluating the overall worth of this research.

The Military Medical Care Environment

After talking with many active duty and retired service members, it appears that it is often difficult for service members to get a medical appointment in the local military medical facility for themselves or, especially, for their dependents. Those currently on active duty seem to complain the least, but their dependent spouses and children, who routinely seek to take advantage of the "free" medical care promised them under the service members' benefits package, have a lower priority than active duty members. The retired military family has implicitly contracted for "free" medical benefits for the rest of their lives through completion of their voluntary obligation of service to country. However, the ever shrinking budget for peace-time military quality-of-life programs conflicts with providing medical care for retirees and military dependents, especially since the wartime readiness and operational support roles of the military medical community take precedence. The obvious result is excess demand on the military health-care system.

The military medical command fears a defacto loss of credibility, and military managers are concerned about the impact on force attrition when medical care and retirement benefits lose appeal. In a June 1988 Air Force Magazine article titled "What's Bugging the Troops", Chief Master

Sergeant of the Air Force James C. Binnicker noted that free medical care has always been seen as a major benefit of service life. Chief Binnicker stated that wartime readiness and peace-time health service is straining the limits of medical resources, and that retirees and dependents are aggravated by lack of access to the military health-care system. He claims they were frustrated by long waits for, and unavailability of, appointments. Regarding the current situation, the Chief attributes this problem to the underfunding of the Fiscal Year 1988 budget for medical programs by \$65 million (7:67). There is concern that the problem of maintaining the health benefits of military retirees and military dependents will be a continuing fiscal problem as Congress decides on how to fund this and various other military quality-of-life programs.

Access to medical care can be viewed as the entire flow of events from the initial attempt to make an appointment, to getting an appointment and travelling to the facility for care, dealing with the people at the hospital, and then, receiving treatment as necessary (34). Each element of this chain of events interacts and has the potential to affect the attainment of a person's ultimate goal of receiving health-care, as well as affecting that person's perception of the quality of care the medical system can provide. Many programs are being put into place to try to ameliorate the problem of accessing the promised medical care. These

programs range from allowing reimbursement and help for seeking civilian medical care to across-the-board audits of efficiency and quality-of-care issues in each military medical facility. This particular study looks at one large medical facility and suggests ways to improve its telephone service with specific attention being paid to callers trying to make appointments at the busiest appointment sites.

Wilford Hall Medical Center (WHMC)

Wilford Hall Medical Center (WHMC) is the largest military medical facility in the southwestern United States. The facility has a 1,009 bed design and employs over 4,100 military and civilian workers, while providing 136 clinical specialties and sub-specialties. It serves a large regional role in providing health care to active duty and retired military personnel and their dependents. In the city of San Antonio are four Air Force bases and one Army post of significant size. Additionally, there is a large retiree community from all the different services in the immediate and surrounding area. Typical service for the medical center's outpatient clinics averages 76,000 visits a month, while true demand is likely to be even higher due to many efforts to obtain appointments being frustrated on the basis of limited resources. While management recognizes it is not possible to expand to absorb the demand, there is strong

sentiment that the public's perception of the quality of care is adversely impacted by competition to get access to the care givers.

System Description of WHMC's Phone System

This study addresses a small, but significant portion of the phones in WHMC. First of all, only commercial lines are of interest; autovon lines and tie lines to the local bases are an insignificant source of traffic into the appointment centers. At present, 90 commercial telephone lines are being leased from Southwestern Bell consisting of the 670- 5000 to 7999 extensions. These lines carry telephone traffic into the WHMC's SL-1 digital electronic switching complex from the downtown switching complex. Any x4000 series extensions at ACD sites were set up internally by WHMC communications staff for intra-WHMC business and are not used for making appointments.

Of the 3086 telephones in WHMC which can access these commercial lines only 21 are used across the ten appointment centers concerned in this study. However, these 21 phones are unique in that they can put several calls on-hold at a time via a call sequencer, the Automatic Call Distribution (ACD) system. With this queuing system and the high demand on these phones during appointment making periods, these 21 phones can account for as much as 10 to 40 percent of all

the incoming calls to WHMC, but as shown in later chapters, seldom more than 12 of these 21 phones are manned even at the busiest hour of the month, and just 6 of these 12 clerks (four different ACD sites) account for the majority of the ACD calls during the busiest hour. Baseline data for this study are shown in Figure 1 which depicts the demand for WHMC and the ACD phones for the first full week of the month of April 1988. The WHMC figures were obtained from the Southwestern Bell Telephone Company representative after a special study, while the ACD figures are the actual data collected from the ACD system at WHMC for the same period.

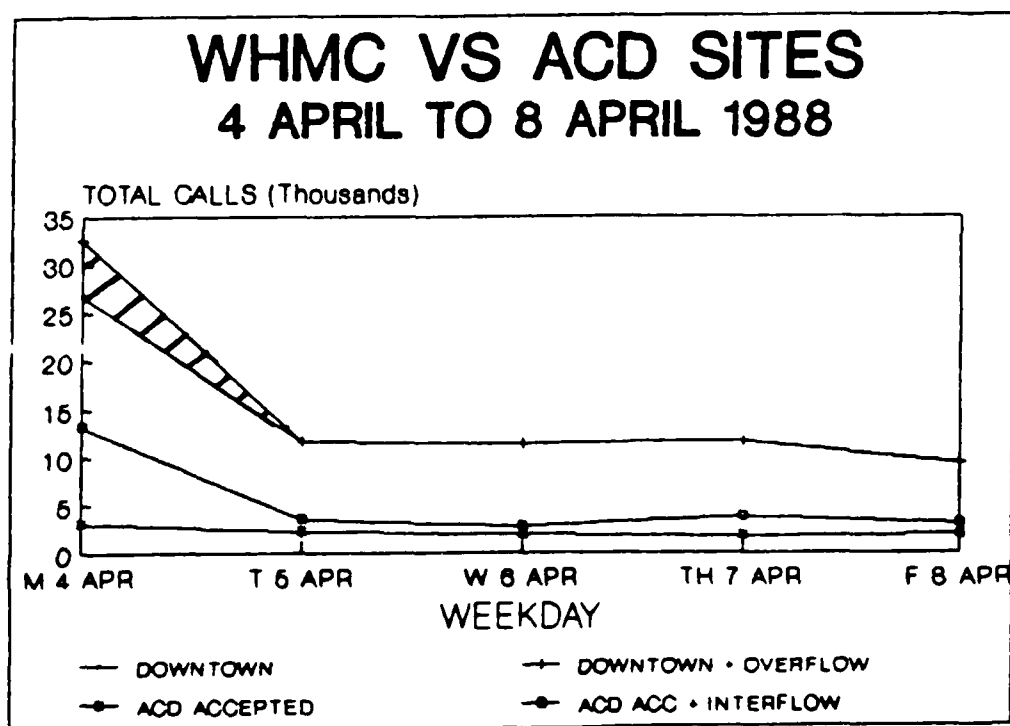


Figure 1. Full System Data

Non-ACD Destination. Figure 2 diagrams the flow of possible traffic over the commercial lines to ACD appointment phones and other phones at WHMC. While a few other sites in WHMC use ACD phones (Central Paging, Biomedical Maintenance, and Patient Escort), they are not for appointments and do not use much, if any, of the commercial trunk resources since they are mainly internal to the WHMC. All traffic not destined for an appointment site has been labeled non-ACD traffic for this study. When the system is congested and a call cannot get a commercial line, but instead the caller hears a fast busy signal, the call is blocked and is referred to as overflow. The caller may then redial the number to try again. When the call gets through it will be routed to the SL-1 switch at WHMC. This switch is configured in such a manner that no blocking will occur across the phones it services so the call proceeds to its destination. If the call is not destined for an ACD site, then it can either be answered, receive a busy signal, be left to ring, or be put on-hold by a typical office phone with two or three extensions so long as the destination phone has this capability. If the call is destined for an ACD site then the logic is more complex.

ACD Appointment Site Destination. Upon arriving to an ACD site several things can happen before an attempt succeeds and the call is answered at the ACD site by an

appointment clerk. If all the waiting space at the ACD site is currently full (queue length presently set at 6), the caller gets an "all circuits busy message" and is labeled as an interflow call.

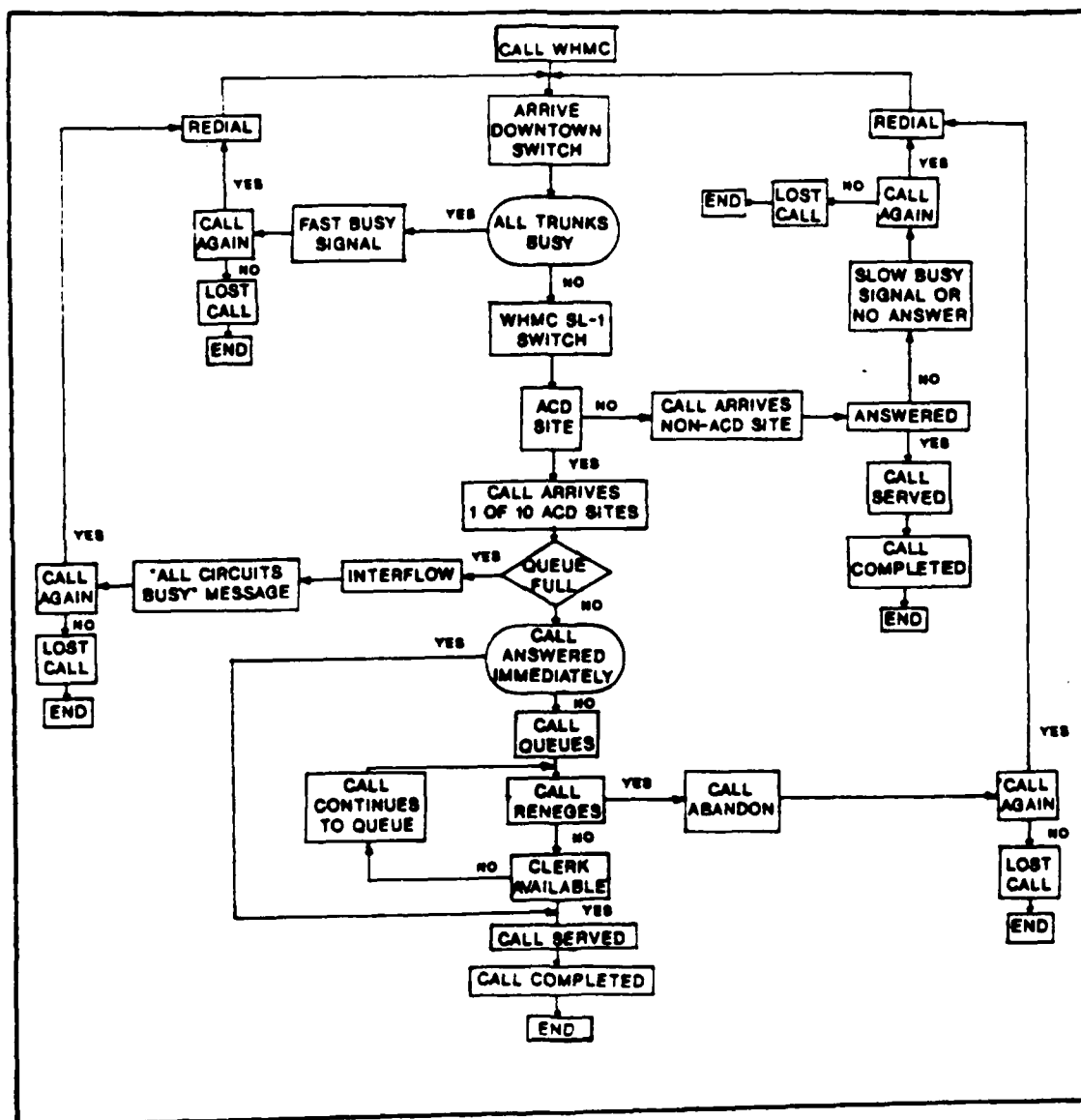


Figure 2. Telephone System Diagram

An interflow call is an attempt virtually blocked at the point of access to the ACD site. If there is waiting room, the caller joins the queue for the appointment clerk. Upon joining the queue (unless immediately answered), the caller hears a recording announcing they have reached a WHMC appointment desk. The recording also asks the caller to be ready with appropriate personal information (social security number, etc.) and encourages the caller to hold-on for the next available server. Music is then played to the waiting callers and if not answered within 30 seconds, a second message is heard which generically thanks the caller for holding, reminds them to have their information ready, and encourages them to hold-on, since calls are answered in the order received. This 30 second pattern of music and the second message continues until either the caller is answered, the caller hangs-up, or for some reason the ACD phone is shut down. If the caller disconnects, the ACD system counts it as an abandonment, or renege, and the average time for all calls abandoning is collected.

The software governing the phones to the ten centers is of unique interest in that it collects hourly and daily queuing statistics of caller and telephone service behavior. These collected statistics reflect many aspects of traffic and caller behavior. Some of the statistics most pertinent to this study are: 1) the number of calls made; 2) the number of calls answered; 3) the number of calls that

received busy recordings or hung-up prematurely; 4) the average time callers waited; 5) the average time a clerk talked to caller; 6) the average time a clerk spent after a call to process information; and, 7) the number of clerks on duty for the hour and their hourly utilization statistics. A key piece of information not given in the ACD data nor tracked by the appointment clerks is the number of callers who actually receive the appointment for which they called and the number of callers turned away.

The Individual ACD Sites. All of the interviews were conducted with the caveat that this effort was not a "witch hunt", so few direct citations will be referenced as supporting documentation for observations made while interviewing personnel at WHMC. During the interview process, the purpose of the thesis was explained as helping the communications staff to choose the best setting for the stack (queue) size across the ACD sites (only one setting can be used for all the sites) and to help them decide how many commercial phone lines are needed. No objective telephone service standard was in place or proposed at the time of this study to show what would be "optimized", but everyone agreed there was a problem with "the phones" and seemed glad to get a chance to be heard.

The ten ACD sites used in this research along with their telephone extensions are listed below. The remainder of this paper will use their acronyms for convenience: 1)

Primary Care, PCARE, 7177; 2) Orthopedics, ORTHP, 7640; 3) Gynecology, GYN, 7785; 4) Internal Medicine, IMED, 7616; 5) Ophthalmology, OPThL, 7841; 6) Obstetrics, OB, 7781; 7) Pediatrics, PEDS, 7515; 8) Dermatology, DERM, 7731; 9) General Surgery, GSURG, 5925; and, 10) Cardiology, CARD, 6534. Note that these are the ACD site names as used by the communications staff at WHMC, and that these sites may also make appointments for several clinics besides the one for which they are named. For example, the Cardiology site gets calls for the Cardiology, the Pulmonary, and Rheumatology clinics. A temporary site for the Urgent Care clinic operated from late May to mid-August 1988, but this site was excluded since it was only a temporary attempt at lowering the traffic into PCARE and better serving calls for urgent care. Traffic for Urology now (since June 1988) goes to an ACD site of its own versus going to GSURG, but not enough data was available for this new site to use it in this study.

Though unauthorized, at least one ACD site which has many out-of-town patients had a 5xxx number which was given to long-distance callers. It served to keep caller long-distance phone bills lower than waiting in the ACD queue as well as serving in effect as a priority line. Published phone numbers and booking times serve as the main source of information for caller awareness to make appointments. However, whenever a special commercial number is set aside

for a special class of callers, a toll free number for long distance callers, or even just an auxiliary number is left on the appointment phone, the calling population finds out this "easy way" of getting through and soon the number is a common point of entrance for all appointment traffic. Paradoxically, much turbulence arises from clinics changing numbers; news of the change does not seem to reach the potential callers, and the clerks pass out the information for these callers (usually with a short explanation of when and why the change took place).

The Process Being Modeled

Reasons for Calling. After interviewing the clerks and their supervisors in each of the ten ACD sites, it was found there are many types of calls arriving at all intensities of caller traffic. A caller contacts an ACD appointment site by dialing a designated number and getting past any downtown switch and queue congestion. ACD sites are intended solely for making appointments, but since reliable information on hospital telephone extensions is hard to obtain, calls may be made to an appointment center for any number of reasons:

- 1) for the prime purpose of trying to get an appointment;
- 2) to confirm an appointment previously made and ask directions for getting to WHMC or a specific clinic;
- 3) to speak with, look for, or leave a message for a doctor;
- 4)

to get a prescription refill; 5) to ask for the results of a lab test; 6) to try to locate lost records; and, 7) for general health information, or a specific question germane to a clinical specialty (e.g., "My cast got wet while showering. What should I do?"). The interviews further revealed part of what the caller goes through, and how the caller reacts when finally reaching the appointment clerk.

Caller/Clerk Dialogue. Once the caller reaches the appointment clerk, a dialog takes place. There are four possible outcomes of the conversation, assuming the caller dialed correctly and knew where to call: 1) the caller gets the appointment sought; 2) the caller cannot get an appointment and is told to call back at the next date for making appointments, or is given information for alternative methods of getting medical care; 3) the caller gets the information desired; 4) the caller cannot be helped at all. Each of these outcomes can each have a bit of annoyance attached for the customer.

There are several elements from the above dialogue that can impact the caller's perception of access to care. In the first case, the appointment received might be two to six weeks in the future, which might be further away than the caller wanted to wait (perceived need is probably connected to perceived seriousness of the medical ailment). In the second case the customer is thwarted and may even have gone

through this call-back cycle before. The third case represents a satisfied customer, but the wait to be answered could be deemed too long. Lastly, no one likes to struggle past busy signals and then have to wait to find out they cannot be helped.

The caller's first impression of WHMC can be affected by other things as well. The above scenario would be exacerbated if the caller was paying for a long-distance phone call. In every case, the caller would have reason to complain if the clerk had been rude. Next, the appointment clerk's side of the phone will be described as a conglomeration of observations across the ten different sites.

Physical Environment and Interruptions. The sites for telephone appointments are decentralized, co-located with their respective clinics. Their construction ranges from isolated rooms, dedicated to the telephone appointment clerks, to desks in plain view of walk-in patients. The isolated clerks are most able to concentrate on the business of answering calls and dealing with callers in all intensities of traffic. Three of the ten ACD sites have this arrangement. Interruptions in their duties for administrative tasks or from passers-by with questions or small talk are kept to a minimum. Clerks at other sites often mentioned being interrupted while trying to answer the phone as a major reason for answering so few calls in the

overload periods compared to the sites without routine interruptions.

Interruption in answering the phone was inherent for all ACD sites. Clerks at every ACD site experienced some sort of interruption: 1) doctors request patient records, their own schedules, messages, or just a social visit; 2) patients ask for directions to a doctor's office or to a lab, or try to make an appointment; 3) the clerks get sent on errands for the doctors to get lab results or records; 4) the clerk had to find a doctor to complete a telephone consultation; and, 5) the clerk had to do other administrative duties besides answering the phone due the clinic being understaffed. Several of the clinics book appointments on Monday mornings when many of these types of interruptions just named are the most severe.

Hardware available at different sites was also a factor in speed of service with potential for improvement. Sites which used a computerized appointment scheduling system were much faster than sites working with sheets of paper in several three-ring binders. The main time saver in the computerized systems was the speed of searching for an available appointment. The computerized systems completed much of the other administrative paperwork such as provider biometrics (statistics of physician activity) and weekly performance statistics, and also kept clerks from trying to use the same book while processing a call. Lastly, the

amount of ACD and computer hardware impact the possible number of agents that can be used at a site. If only two appointment phones or two computer terminals are physically present at a site, then a third server would only be useful to relieve another clerk versus being fully operational. Therefore, hardware must be allocated for use by part-time standby servers in peak conditions, but the equipment may be unused much of the rest of the week.

Tasks and Responsibilities: A Function of Staffing.

The depth in numbers of administrative personnel has much to do with the amount of interruption allowed in the telephone appointment task. Clinics short in staff depend on the clerk for a wide range of tasks which are often needed at the same time as phone traffic is high. Manpower is hard to get, and most clinics depend on the clerk to do a variety of other tasks beyond answering the phones even though that is the task done by some clerks most of the time. A consideration about tasks done in the work environment is career progression of the clerks. The 906XX, Medical Administrative Specialist, career field is much broader than answering the phone for whatever purpose, but it is often apparently the case that clerks are left in these jobs for long periods of time. While it is clear that this is a motivational problem, it is only speculation at this point as to how this lack of experience impacts these 906XXs' ability to compete for promotion in the Weighted Airman

Promotion System (WAPS) testing which evaluates the clerks career achievement.

ACD Site Supervision. Each clinic had its own philosophy on how to make appointments and how to utilize their clerks. Note, that "policy" was not mentioned as no stated policy was found in any site for answering the phones. There was a marked absence of priorities, procedures, guidelines, or protocols for the use of the clinics' telephone appointment resources. Forecasting telephone demand and planning staffing requirements has never been done, but clerks know the busiest times of their site. If operator service were increased it would only serve to expend all available appointments just that much faster, leaving even more time for the clerks to turn away callers. Being able to shorten times-on-hold for those getting appointments, or cutting down on the number of busy signals callers received seemed a waste of energy to many clerks interviewed given the ultimate result of the extra effort.

The Weekly Experience. In every clinic interviewed, the clerks felt distress from the activities involved in answering the phone and in dealing with patients. It was made clear that answering the appointment phone as a full-time job was the worst administrative job to have in the entire medical center (working in the records section was a close second). Monotony, radically changing workload

(overload to slow with spurts of activity), and having to talk with frustrated people were the chief drawbacks. The ACD hardware uses a headset which saves the clerk the cumbersome telephone, but which also enslaves the clerk to the immediate area. This works well for the overload periods, but some clerks felt trapped for the other 60 to 80 percent of the time in the week. The most common observation is that the number of available appointments is much less than the number of people calling for appointments.

This excess demand often creates a situation where the clerk must turn away callers until the next opening of appointments, when the caller must then compete in an "open lottery", so to speak, to try to call in before everyone else to get an appointment. The frustration in the clerk being unable to help and in the callers being thwarted from their goal (possibly over and over again) can be seen as an aggravating and chronic problem for the clerks. This pattern of distress varies by clinic depending on their policy for opening appointments, the demand for that clinic, and possibly, on the perceived seriousness of the need by the caller. This "turn-away syndrome" leads clerks to ask why bother to man the phones to talk with frustrated callers, whom the clerks cannot help, when the clerks believe they could use their time more effectively. This

syndrome was the chief source of stress labeled to incite rudeness in both clerks and callers.

There are other items that tend to stress the clerks that do not occur in a weekly pattern. When the computer system goes down, no appointments can be made, and if it is an appointment-making day, the clerks must face the overload of calls at least twice, since callers are typically asked to call back that afternoon when the computer is expected to be working. Most clerks would like to have a message machine to handle the calls during this type of event, but communications hardware cannot presently support such a request. Callers who have hearing or speech impediments, or do not speak English well tend to stress the clerks, as do callers who want to make small talk with the clerk during overload periods, instead of just giving the necessary information to get an appointment.

Clerk Distress: Management. A further source of distress for the clerk is their management. The foremost complaint about management was the lack of any ACD site-specific standards for the service of the phones to use as feedback. The drill seemed to be book appointments until full and then deal with the rest of the callers, with an occasional cancellation opening in time to help a caller. The next biggest complaint was doctors changing their appointment schedules on short notice forcing the clerks to cancel and reschedule patients. The clerks felt the doctors

did not pay attention to their schedules beyond the next few days, while the appointment books run out for two to six weeks. The doctors often had many changing demands on their time which forced them to cancel appointments on short notice. This situation often led the patient to expect too much in terms of how soon they would be seen when calling to make a return visit appointment. Throughout the interview process it was observed, with a rare exception, that there was a perceived lack of management supervision and concern in regards to the appointment clerks' duties.

The Ambulatory Care Initiative: WHMC and Access of Care

Management at Wilford Hall Medical Center (WHMC) has responded to today's demand situation like many other military medical service organizations. The WHMC Commander, Major General (Dr.) Vernon Chong, made access to health care services his highest goal second to the enhancement of Wilford Hall's operational support mission and medical readiness. There has been much effort to enhance the access situation by implementing programs aimed at increasing staff, streamlining procedures, and many other management initiatives. In 1987, the Air Force Medical Service began a program titled the Ambulatory Care Initiative which is a broad-sweeping program aimed at improving the accessibility (or the perception thereof) of military medical facilities

to the patient population. This initiative encompasses the spectrum of medical facility activities to include telephone access to medical facilities and their appointment systems. This initiative coupled with several chronic phone service complaints of October 1987 was the basis for the studies branch of Wilford Hall (WHMC/SGMB) to sponsor this thesis effort.

Documented Telephone Access Problems. The Patient Affairs Office conducted a study of its outpatient survey data from October to November 1987 with some interesting results. There were three main indicators of dissatisfaction and ten main areas of concern brought to light in the summary. The three main indicators were: 1) courtesy; 2) time to reach an appointment clerk; and, 3) the reasonableness of the appointment time (how long to wait to get in). Senior staff in the Patient Affairs and Quality Assurance offices of WHMC have acknowledged that most complaints arise from rudeness or unprofessionalism perceived by the caller, rather than from just not getting an appointment. Thus, there is potential to affect the perceptions the callers have of the quality of health care the hospital (or the military medical community in general) is providing at the time the caller speaks with an appointment clerk. Time to reach an appointment clerk by phone is an item often marked as poor on surveys of outpatients, as is time to get an appointment. These two

items may have confounding effects for people trying to make contact solely by phone. It may take several tries and a long wait to get to talk with an appointment clerk with no guarantee of getting an appointment. It is possible to fail to reach a clerk in time to get an appointment week after week. The time until being seen once an appointment is given out varies by clinic from 2 to 6 weeks from the time an appointment is made.

The ten different areas of concern can be classified into action issues of clerk training and staffing, phone system configuration, doctor management, and excess demand. The areas of concern that apply to this research are ranked in order of complaints most often received: 1) the caller could not get through to an appointment clerk, but kept getting a busy signal or a recording stating "all circuits busy", so they would redial many times, and yet they had no idea of appointment availability; 2) high phone bills of long-distance callers who had to wait in queue; 3) the length of wait to be seen; 4) the lack of available appointments -- getting through to the appointment clerk, and then being told all appointments were full, so call back (next week) when more appointments would be made available; 5) the length of time on hold; 6) appointment phones being shut down while callers were still waiting in the queue (by accident or at the close of day); 7) a return visit cannot be made as the doctor instructed nor can it be made while

the patient is at a clinic after seeing the Doctor in a clinic which has an appointment by phone only policy; 8) the clerk appointed them to the wrong clinic; 9) the appointment was canceled on too short notice and worse, on an out-of-town patient who had travelled to WHMC; and, 10) an appointment was made for one doctor and a different doctor actually saw the patient. This research is aimed at the first two action issues identified above.

Changes in the System Over the Course of the Study

This study was conceived in October 1987 when management at WHMC was just beginning to deal with the onset of the Ambulatory Care Initiative. Management was not sure exactly what was wrong with the system, but several outstanding areas of complaints demanded attention and action. The three milestones important to this study occurred in November 1987, March 1988, and around June of 1988.

The first squeaky wheel was a large number of long distance callers getting high telephone bills from long times on-hold (some as high as 45 minutes). These people became angry at the cost as well as at the experience of holding for long periods and then either not getting the appointment for which they called or being disconnected by accident or by the appointment desk shutting down for the

day. Many clerks and even the Patient Representative recall having patients presenting their long distance phone bills to them upon coming in for their appointment. In November 1987, the communications staff changed the policy for stacking at ACD sites from virtually infinite stacking to stacking to only six per site. Six was picked with the intuition that if most calls were handled in less than two minutes then the worst case for average waiting would be about ten minutes. The effect of the policy change was that it did cut down on the complaints of long wait times and also it helped cut down the most severe phone bills.

The second major change in the system was the addition of 30 commercial lines at the end of March 1988. There was money available and the intent was to reduce the congestion throughout the WHMC system by adding more lines. Though little is known about the congestion with only 60 commercial lines, it can be seen now that the system only experiences blocking due to congestion for 2 to 3 hours on Mondays.

The third change was the realization of the Ambulatory Care Initiative in the showcase setting of the ACD site at the Primary Care clinic. The Ambulatory Care Initiative personnel have authored an operating instruction (OI) for ACD site telephone service which allows for each clinic to tailor the guidance to its own situation. The OI was being put into place in July of 1988. The new capstone policy calls for attention to be paid to answering the phones, for

courtesy to the caller, and for standard training for all appointment clerks. The OI sets management responsibility for the behavior of the clerks where lines of authority and priority have long been obscure.

Management Update

Currently, the two "telephone" elements of the problem are now perceived to be handling the surges of demand, and determining the best configuration of the number of city trunk lines to lease (90 lines now) and the stacking policy to set across the ACD sites (stack to 6 now).

The crisis of September 1987 left a legacy of management initiative that has been growing and gaining momentum. The underlying problem of patient (perception of) care seems Air Force-wide in scope and can be seen in the Medical Service-wide Ambulatory Care Initiative. As mentioned above, the doctrine calls for empowering a high-level management team to cut across all existing organizational lines in order to improve the efficiency and sensibleness of operational policy and resource (capital and human) utilization for improving access to care. Among other things, the initiative sets out broad policy and goals for care providers and for support staff in many areas of patient access to care. The showcase for the Ambulatory

Care Initiative has been the Primary Care Clinic, as already mentioned.

A partial result of these efforts has been evident in WHMC by the effort to create a training program for appointment clerks, and to consolidate the training for appointment clerks across the hospital. Lessons in courtesy, problem resolution, stress management, and the "big picture" are aimed at achieving a consistent high standard of military professionalism, personableness, and helpfulness in each of the clerks. Lessons in using the ACD hardware and in using a particular appointment log system (manual or computerized) are aimed at the working efficiency of the clerks. Policy is set so that the objective of the phone dialogue is to make an appointment as quickly as possible, without a lot of small talk, in order to increase the number of calls that can be handled during the periods of heaviest traffic.

Organization of ACD Sites. The current configuration of the ACD sites is decentralized. The communications staff wants to gain efficiency by consolidating. The doctors want to keep the status quo to keep the clerk manpower available to them for answering the phones and other duties, some of which are clinic specific in nature. The senior administrators do not want to go wholly centralized, but they want to get more efficiency out of the clerks. They have argued for a "cluster approach" (34) which would

combine the ten ACD sites into three or four sites. They hope to save manpower and see some improvement in telephone access to the WHMC. The clerks are willing to work, but they do not like answering the phones when they cannot help the caller. The clerks also seek guidance for what telephone answering standards managers want, and get frustrated over shifting task priorities. None of these groups have yet to propose a measurable standard by which to consistently gauge telephone access and evaluate caller service. The OI does give some indication of using ACD data as feedback to the ACD sites for the average time on hold of callers, but a measure of access is still nonexistent.

Hypotheses for this Research from the Background Investigation

Across the board, all the clinics make appointments and fill the books for the clinics. The most efficient telephone operators only succeed in filling the available slots in their clinic's log faster than the less efficient operators. The remainder of the callers then participate in a dialogue to alleviate their feelings of not getting into the system with a recommendation of when it is best to try calling again, or to get advised of their alternatives for finding care (emergency room, Health Care Finders, CHAMPUS, etc.). Depending on the number of appointments available

and the demand at the time, some clerks have estimated that in a week every caller they talked with after Monday afternoon was turned away for a regular appointment due to a full appointment log. During periods of slow traffic, especially after all appointments are gone for the week, most clerks feel their time could be better spent doing other duties away from the phone. It is far beyond the reach of this research to create more appointments; the staff at WHMC and the guidelines of the ACI both work well to create the most appointments physically possible in the situation of excess demand and priority classes of appointments. The best result that can be hoped for from this research is that all callers to WHMC will perceive getting better service (quickly obtaining what the caller wants), and that WHMC can find a way to use its personnel and telephone hardware more efficiently.

Demand. In an interview during July at the Primary Care Clinic, Dr. Fred Kron characterized demand for care when he said, "Wilford Hall is like a huge HMO, except we don't know the extent of our subscriber base...who have contracted for their health care benefits by 20 years of service." In other words, the true demand for service at WHMC is an unknown. Further, when it comes to telephone demand, no long term record was in place to forecast patterns in caller behavior; however, one is currently being implemented. The communications staff which monitors ACD

system software has recently begun using a spreadsheet to store a database of ACD call data. In the long run this information should reveal the daily, weekly, monthly, and yearly seasonality of telephone traffic demand to the various ACD appointment sites. This data can help to forecast demand for better use of augmented manning for ACD sites, identify peaks of excess demand for examination to see if policy changes can lessen the peaks, and track the actual effects of policy changes on caller behavior.

Scheduling. Across the ten ACD sites under study, the scheduling of clerks against demand patterns ranges from the best that can be done to an atmosphere of unconcern. The best scheduling was found in the showcase of the Ambulatory Care Initiative, the Primary Care Clinic, where clerks were staffed against peaks in demand. In several clinics the clerks stated they tried to work together on the peak loading periods while keeping interruptions to a minimum during the peak demand periods. In several other clinics the clerks were so overworked they could not handle the phones while tending to other duties. In one clinic the supervisor set the priority of the clerks to administration while they would interrupt this work to answer the appointment phones periodically. In two clinics, clerks who were unsupervised were idle and no one was manning the appointment phones even though calls were queuing.

Feedback. In no clinic, except Primary Care, was there evidence of tracking the actual performance of the appointment phones via the data from the ACD system. The data on ACD site operator performance was readily available and was also being presented in summary form to high levels in the hospital administration, but ACD site managers did not routinely use the information for determining telephone access to the clinic. However, the raw data from the ACD can be difficult to interpret as the example in Appendix A shows. Findings from this research should help to format the information in a more useful report for senior management review and ACD site feedback.

What is Needed. The ACD sites need standards they can support which can be published for callers in order to allow customers to infer what to expect (long wait, busy signals, time until getting an appointment, etc.) when calling an ACD site. Efforts should be made to get the most people the best service possible, while trying to limit the telephone hardware costs and getting the most efficient work out of the clerks. The organization openly admits it cannot handle the excess demand and is trying to find other ways to get patients medical care. One hypothesis is that forecasting telephone traffic and appointment availability will allow informing callers of what to expect when attempting to make an appointment. Another hypothesis, is that lessening traffic on Mondays should allow normal business calls freer

use of the telephone resources. Standardizing service times at the sites via supervision and ACD feedback, coordinating scheduling of booking days, limiting the queue size to stop callers from renegeing, and increasing clerk productivity are other areas for hypothesis and analysis. A simulation model of the system was created to gain insight into the telephone system behavior, and to evaluate hypotheses on the system.

III. Literature Review

Literature Review

The greatest effort of the review was to try to understand the true nature of telephone systems and caller behavior, while staying close to the context of the operational problem at WHMC. The body of this section will discuss the relevant issues found in the literature reviewed to date in the categories of telephone systems and telephone traffic, simulation of telephone traffic systems, queuing theory, and caller service.

Telephone Systems and Traffic. The telephone system at WHMC, at first glance, appears to be similar to many systems found in industry. Patients phone the hospital and either get a busy signal, get a recording saying "all circuits are busy", are put on hold, or their call is answered. A busy signal results when all telephone lines are busy. The "all circuits busy" message is given when the call is blocked from joining the queue of calls on hold. Being on hold equates to waiting in line (queuing), and calls are answered in the order in which they are received (first-come-first-served). A caller on-hold may decide to hang-up (renege). If answered, the patient may or may not get an appointment. The pool of people who got a busy signal, a recorded

message, reneged, or did not get what they wanted may try again at another time (retrials). Articles of seemingly similar systems, and articles analyzing the phone traffic in them, are numerous in the literature.

There are several journal articles which deal with systems comparable to WHMC. The most pertinent of these deal with the following systems: 1) the IRS taxpayer information system (26); 2) an airline reservation office (44); 3) an insurance company claims and policy division (23); 4) a Bell Canada business office (49); 5) a New York bus terminal information desk (12); 6) an Australian betting agency (66); 7) a Bell System direct dialing study (45); 8) a Bell System operator staffing study (63); and, 8) a child abuse hot line (46). Review of these and other articles demonstrated a common problem and gave insights into how to approach WHMC's problem.

The common problem identified in all the above mentioned articles was a trade-off between level of service to the customer and the cost of manpower and equipment of the system. Dickens (12:4-5) illustrated the problem clearly in a simple example, where 36 clerks were needed to answer 100% of 36 calls to maximize service, or to have a much smaller number of agents work 100% of the time to get full utilization out of them. Somewhere between these two extremes, a level of service, or instead, a desired efficiency of overhead, is chosen by management either by

rule of thumb or empirically in all cases (12:4; 44:3; 23:105; 26:505; 33:208; 46:21; 49:159; 60:1; 66:991). In the attempt to answer this problem a few common themes arose in the literature.

There are four main points of interest in the literature on the trade-off problem: 1) determining the actual demand for service; 2) determining system size and service parameters; 3) modeling the telephone system; and, 4) implementing study findings. Each point will be discussed in the order listed.

Determining Telephone traffic Demand. The actual demand for service varies across time and with certain types of system and caller behavior. Intensive data analysis is required to characterize the number of calls made to a system across the day, week, month, or time of year. The total demand is often visualized as a random variable with a time-variant Poisson distribution whose parameter changes based on busy times during the day, week, month, or season (12:6-11; 23:106; 26:511-516; 33:210-212; 35:1288-1300; 38:125-133; 45:298-299; 46:22-23; 60:1-2; 66:992-993; 63:223). The demand with overflow to a trunk group has usually been modeled as an interrupted poisson process (35:1295-1297; 21:557-562; 27:1225). The demand distribution is affected by both the system configuration and customer demand.

System and caller behavior can cause variability in demand so that the total measured attempts to reach a system overestimates the true demand. This occurs when callers retry a call after their call was blocked due to a busy system or when they reneged after being put on-hold. Falin (16:37) and Elldin (14:342) both identify that typically, blocked callers will immediately retry their calls until convinced they cannot get through. Methods which attempt to explain and try to reduce these high estimates have been proposed (21:558-568; 26:515-516; 33:209-213; 35:1284-1300; 44:12; 38:125-135; 39:1128-1137; 45:295-311; 60:2; 63:233; 14:342; 16:40-41). Liu (45) offers actual telephone traffic data showing retrial behavior while the rest of the articles hypothesize their own models for dealing with retrials, or simply ignore retrials to simplify their studies. Analytic methods for dealing with retrials exist, but are very difficult to model or obtain numerically (21:558-563; 39:1126-1128). Hoffman, Harris, and Saunders (26:513) set the retrial section of their simulation model to zero, since so little was known of the actual retrial distribution and since they observed that retrials only affect demand and not the rest of the process.

Most authors saw the need for long term tracking of the actual demand patterns for use in forecasting demand and validating work force levels. Sharma (60) suggested a coarse estimate could be derived through observation to

build a "Time Consistent Daily Profile" across the days of a month. Sharma also observed that if the demand had different destinations as in an airline's system (reservations, baggage, etc.), then data would be needed for each type of activity. There were many authors who warned of the effects of variation of traffic levels when accounting for dimensioning telephone circuits, especially for daily demand peaks (morning and afternoon) and yearly peaks (around holidays) which could either result in excess call blocking or over-provisioning of trunk lines (14:319; 29:561-562; 50:967-969).

Since the blocking in a telephone system is so sensitive to variations in traffic due to peakedness and retrials, traffic studies often focus on the single hour in which the system is most affected when defining the trunk size required (4:2352). Neal and Kuczura show that at least 20 single-hour measurements are needed to estimate call-congestion, trunk usage, peakedness, and the average offered load to be able to accurately determine the number of final trunks (those going to a specific destination) needed (50:984-985). Jagerman (35:1302-1309) gives easy to run computer programs to allow traffic calculations and even a simple queuing model with delay based on measures of traffic overflow and peakedness, but the apparent simplicity of the programs are confounded if appropriate measures are not taken to gather enough data to make accurate predictions.

Once an estimate of demand has been determined, the ability to serve this demand must be considered. Now, the interplay between final trunk group sizing and the number of operators used becomes the driving factor in service to callers and in cost efficiency for the organization (32:20-26).

Telephone System Sizing. The size (number of lines) and servicing (number of operators) of the system has a great deal to do with the arriving stream of calls, the level of service provided to the customer, and the cost of the system (64:396-399; 35:1283; 20:805-807; 50:967-969; 47:2576-2580). If for a given level of demand (call arrival) the number of lines into a system is too small or there are too few servers, then there will be a high rate of busy signals. If calls can be put on-hold, but there are too few servers, there will be a long wait for service resulting in greater caller reneging behavior. On the other hand, it costs money to buy and maintain telephone lines as well as to hire telephone operators. Methods were identified in the literature for sizing a system based on demand and blocking probability, and for measuring service times of calls. These two topics will be discussed below in turn.

Choosing the size (number of telephone trunk lines) of the system is very important for adequately serving a varying demand pattern, while keeping the number of calls blocked to a low level. System blocking results in a lower

level of service to the caller, contributes to the overestimation of demand, and was found to be sensitive to peaks and variations in demand traffic (27:1215-1216; 35:1297-1298; 38:123; 45:303-306; 64:397; 14:321-328; 38:123-125; 47:2575; 29:561-568). Blocking in an organization's telephone system is referred to as overflow, which results when the lines available are flooded with calls, and incoming calls cannot be connected. Customers get busy signals, and as they retry their calls the true demand becomes inflated. Fredericks and Reisner demonstrate that for a one percent blocking system with modest retrials, the two server case has a retrial blocking of 51 percent (21:569) which they postulate might have an affect on customer satisfaction (or annoyance). In a study for locating answering sites for a non-competitive, public sector service agency, Weston found blocking percentages as high as (2-20%) during peaks in demand (64:397).

The chance of overflow occurring is called the blocking probability for the system, and this measure along with the estimated demand on the system are used to determine the number of telephone lines needed in a system (35:1290-1291; 38:125-130; 14:335; 47:2576-2582; 65:602-604; 21:558-561). Typically, the commercial telephone systems are engineered for an average blocking target of 1% in the busy hour of the busy season (38:123; 61:40; 50:968). Unless the day-to-day load variation of demand is very high, the current method of

calculating the number of trunk lines is quite adequate (35:1297-1298; 38:123). Once a call gets into the system, the time it takes to serve it plays an important role in the performance of the system. Duffy and Mercer (18) suggest that calls which cannot be completed have a significant impact on the total call attempt time as well as on the traffic load on the network. They recommend encouraging callers which cannot complete their calls to abandon their attempts faster thereby releasing trunk (and network) resources more promptly in order to reduce the load on the network (18:33).

Telephone System Servicing. Measuring the time to do the tasks of answering the phone and completing post-call work was done for several of the systems in the literature, and several probability distributions for the service times were suggested. Direct call processing time and post-call processing time for calls are important factors in caller time on-hold, call throughput, and system overflow due to time congestion (27:1221-1225). Some authors suggested combining the two times and using an exponentially distributed random variable (12:12-15; 26:511; 33:208; 39:1131-1137; 46:22; 49:159; 60:2). Jolley and R. J. Harris (37) devoted an entire article to the question of combining the two times or not. They found it permissible to combine the times so long as the post-call-processing time was not too large in relation to the call time (about $< 50\%$ is

satisfactory). This idea was also stated by C. M. Harris and others (26) who found the times to be Weibull distributions (with shape parameter about equal to 1) that could well be approximated using an exponential distribution for simplicity.

Sze (63) noted that the service time for the large number of Bell System operators was not exponential at all, but was instead bi-modal, with the operators working faster at high demand times and slower at other times. Others suggested using an Erlang distribution (23:106), or modeling the service time based on the type of service being requested (66:992). The telephone industry routinely uses the Erlang B and C formulas to estimate the number of lines and servers needed for achieving a given level of service. However, many authors conclude that the assumptions for using these formulas are violated in systems with high degrees of peakedness, losses, and delays (as is the case for service systems which cannot afford to operate at a level where few calls get busy signals, or where renegeing and redialing are commonplace) (32:20-30; 11:383; 27:1215; 65:605). Once the system demand has been characterized, and the system size and system service time parameters have been initially set, modeling of the phone system can begin.

Modeling Telephone Systems. Two modeling and analysis techniques from Operations Research are particularly appropriate for the topic of queuing in

telephone systems. These techniques try to relate the interaction between efficiency and level of service for a telephone system (23:108). For simple systems (general queuing, markov process, etc.), an analytic approach is appropriate, but for more complex systems (queuing with blocking, reneging, and retrials), simulation is often used to gain insight for controlling the system; however, there are instances where analytic approaches have also been applied to the more complex system problems. Once the system was configured, linear programming, integer linear programming, and network analysis techniques were useful for scheduling. The literature provided examples of each type of approach.

There were three articles uncovered in the literature which dealt directly with the simulation of phone systems. The first of these to be discussed was by Globerson (23) who used GPSS over a much simplified queuing structure to analyze operations at an insurance company's telephone claims office. He showed the usefulness of simulation by combining work centers to examine the change in efficiency. He observed that so long as there are no overriding costs in cross training operators, it is more efficient to consolidate phone divisions (23:110). Rothkopf and Rech present a case where servers have specialized tasks, and the cost of cross training makes combining queues a

counterproductive effort rather than an efficient use of manpower and training (55:907).

The second simulation article for review was by Linder (44) who discussed the value of simulation in a study of coordinating services for an airline telephone reservation system. He also discussed the study's implementation over a five year period. Linder's article developed the idea that simulation can give good results on the behavior of phone systems, but that the form of presenting those results for managers to use could impede the usefulness of the findings (the tables constructed in the initial study for finding the right number of lines and servers were difficult for office managers to interpret).

The last article found on simulation of phone systems was by Harris, Hoffman, and Saunders (HHS) (26). The HHS model was written in Simscript II.5 and was much more ambitious than Globerson's and Linder's. The HHS model included blocking, reneging, and retrial routines to mimic the actual behavior of the IRS phone system. However, the HHS effort did relent when not able to estimate all the parameters desired. Reneging and retrials arose as sensitive variables in the model equations, but were essentially assumed out of the simulation model by assuming a constant renege rate and entering zero values for the redialing parameters (26:510-513). Follow on work to the

HHS study led to a solvable analytic model that will be referenced with the other analytic models below.

The literature search yielded several analytic models on telephone systems. The telephone system models were of two forms, queuing models (12; 46; 66; 65) and Markov processes (33; 39; 49; 21). The queuing models either disregarded complex demand traffic (12; 66; 65), or had clearly defined levels of service which did not permit significant losses or delays (46). The literature on queuing theory revealed many special case models with confining assumptions (arrival process, queue length, number of servers, etc.) that were not of direct use, but the queuing literature did help to describe the relationships between the arrival stream, the number of trunks, the number of servers, and the service rate as well as to define the limitations of this analytic method (2; 59; 22; 36; 62; 24; 15; 25; 48; 1; 40; 43; 18). The telephone system models based on Markov processes went far towards trying to incorporate blocking, reneging, and retrials (33; 39; 65; 21), and one even considered the use of stand-by servers (49). The HHS simulation study (26) was succeeded by Hoffman and Harris (33) who developed a Markovian model of "a multiserver loss/delay queue with retrial and reneging" (26:504). All of the modeling efforts set varying levels of customer service and then evaluated system efficiency to show decision makers the alternatives, but it came down to

the problem of implementation of operator scheduling and staffing policies to finally solve the problems studied.

Implementation Issues for Phone Systems. Operator scheduling and staffing for a given level of service must vary with hourly, daily, weekly, and seasonal demand traffic fluctuations within the constraints of labor policy and total budget (12:15-27; 23:108-111; 26:520-522; 33:209-213; 46:23-24; 49:162-163; 51:55; 66:993-994; 65:8-15). Most articles provide tables or graphs for management use in determining the number of operators to staff for a particular time. One article of scheduling operators explores integer programming (51), while another compares and contrasts the efficiency and practicality of linear programming and a network flow algorithm (66). Dickens study (12) ended by presenting a simple heuristic scheduling system including break times, relief periods, and replacement policy. Mok and Shanthikumar (49) created a computerized system to forecast the optimal number of full-time and stand-by servers to cost effectively achieve a desired service level. If maximizing system throughput is an objective, pooling servers as previously mentioned is a step toward efficiency, but when pooling is not possible it is proven that the best enhancement to throughput is achieved by allocating additional servers to the sites with the highest workloads (59:333).

A part of any implementation are measurements of performance. While some specific examples of standards of service are below, Fortuin (17) gives guidelines for implementing and using performance indicators. Fortuin sees performance indicators as a means to an end, and he warns that indicators are only sensible if continuous improvement is desired in the area being measured (17:6). If this is desired, he adds (17:6) that the program must: 1) have top management support; 2) be installed with the help of the users; 3) have quick feedback of results, and management should reward progress, if only by taking an interest; and 4) the parameters in the performance indicator must be controllable by the users, and also be achievable.

Standards of Service in Telephone Systems

Service standards can be of the qualitative or quantitative nature, but are often actually a combination of both (44:3). As mentioned earlier, the telephone industry often uses the quantitative measure of less than one percent blocking in the trunking system as the desired service level for engineering the size of trunking groups. However, when the economy of scale and vast communications network of the telephone industry are compared with customer service at a particular business site the literature shows different policies gaining favor. It is clear that the type of

business has a lot to do with what degree of service is deemed appropriate.

The quantitative measures of service vary widely and systems constrained by budgetary ceilings often have to make hard choices. The Bell System operator system was designed to answer quickly enough that no more than ten percent of the callers wait more than seven seconds to be answered (63:230). Sharma (60) related that the transportation industry (airlines, busses, etc.) often use a grade of service factor such that 85 percent of the calls are answered in less than 20 seconds. While, the Child Abuse Hotline case study (46) showed a low tolerance for missing a call due to busy circuits and the desired blocking probability was much lower than that for which the telephone industry engineers its lines. Linder (44) would determine the number of lines and staff for the airline reservation offices using management prescribed upper limits on the percentage of calls blocked and time on-hold. On the other hand, Hoffman and Saunders (32:28) show the management of the IRS phone service making a conscious decision to have callers get busy signals rather than to have long times on-hold when the system is congested. The IRS management then defined acceptable service based on an hourly objective of waits between 20 and 40 seconds with less than five percent reneging observed (32:28). On the other hand, Kolesar (41:16) suggests using the percentage of customers lost to

the queue to be more important than the actual wait in line in a study of automatic teller machines. These standards allow decisions to be made on the affordable levels of hardware and staff resources, but standards for the qualitative aspects of service can also be important.

Perspective and psychology are two factors often thought of when the issue of quality is raised, and Larson (42) focuses on these two factors in an article on perceptions involved in queuing. Larson's main point is that in everyday life, the psychological experience of the wait in queue can be more important than the actual delay. Larson describes the term "social injustice" (42:896) as a perception that a queue is not fair if it allows people to pass, or be passed by, someone else. He gives several examples of customers being more satisfied when they believe a system to be first-in-first-out, than if everyone's average time would be shorter using a different policy. He links the empty time of waiting in queue with a form of imprisonment with no definite termination (42:897), and he describes how several businesses succeed in filling waiting customers' delay time with music, information, television, or even live performers so their perception of the delay is more positive. Larson emphasizes that the use of feedback (42:900) to allow a customer to estimate their wait makes them "feel better" about having to queue. Larson's bottom line was that the elements of social justice, waiting

atmosphere, and feedback each have a key role in a customer's perception of the queue which often can override the actual delay (42:901), and that reducing the negative side of waiting can often be more cost effective than adding servers or technology (42:904).

Little else of the literature on quality of service was investigated due to the quantitative nature of this study, but there was a particularly insightful article on customer service in industry (e.g., airlines, banking, retailing, electronics, automotive) reported by Patricia Sellers, a reporter associate, in Fortune magazine (57). Though a secondary source, the information presented could be useful when planning to implement any change in service. The article described how handling customer complaints could affect a business's success as well as the customer's perception of that business (57:87). Linder recognized this in his work with airline reservations by noting that the telephone was often the first direct contact with a potential customer (44:3), but his pursuit was quantitative. Sellers reports from interviews that since customer needs often dictate the market for profit-making industries, customer complaints and the way in which they are handled are an important part of successful competition in the open market (57:92).

Sellers reveals several methods used in industry to handle customer complaints. It was noted that the hardest

part was getting the customer to complain to the company, since customers are likely to tell twice as many people of bad service as they are of good service (57:87). She reports that simply listening to the complaint can increase brand loyalty, and that often the customer can tolerate a delay more comfortably if given a full explanation for noncompliance with their needs (as in a delayed airline flight) (57:89). The need for companies to train, monitor, and motivate their telephone operators is apparent since being rude or saying "It's our policy..." to a caller with a complaint is very negatively received -- one company is said to evaluate their operators on productivity, attendance, attitude, and quality of service (57:96). A bank president is reported to personally answer a small percentage of the mail-in complaints to show that top management is concerned and involved with customer problems (57:92). In other companies, supervisors operate under the policy of "serve them well from the start", and they remind themselves and their operators of service standards by keeping a list of the five worst things they could do to a customer (and ways to avoid them) at their desks (57:100). Lastly, Sellers cited that reliability is ranked high as a factor in good service, and that when advertising, care should be taken to not promise more than they can deliver (57:100).

Discussion of the Literature

The literature revealed a great deal about the complexities involved in telephone systems and their modeling. Overall, telephone lines are the lifeline of a business and servicing customer demand is of foremost interest. While operating at 100 percent service is not practicable, most organization's management set an objective level for service and then procure however many lines and operators necessary to fulfill that objective. When customer demand varies, the number of operators are typically increased, or decreased, accordingly as hardware and personnel policy allow. Much research is aimed at figuring out how to provide the right number of servers and lines to meet service objectives of caller blocking and time on-hold in the most cost efficient manner. The published research uncovered to date indicate that the Erlang formulas B and C, queuing models, Markov processes, and simulation have all been applied to different settings in the area of telephone traffic and caller servicing. A discussion of the impact of these findings on the current study of WHMC's telephone system and servicing follows.

Why Wilford Hall is Different. As with the studies of the IRS and the Child Abuse Hotline cited above, WHMC is a not-for-profit operation; however, WHMC is distinguished in that it has many different streams of incoming calls to be

serviced by queuing at decentralized sites, and manpower is limited. Furthermore, while the office phones throughout the medical center have a weekly peak usage on Monday (especially in the morning), the ten ACD appointment centers in this study each have their own peaked demand periods across the week and within the month (with the worst aggregation of these occurring on Monday morning of the first full workweek in the month). This busy-time WHMC telephone system could be described as a two stage, heterogeneous, multi-site, multi-server, loss/delay queuing system with redialing. WHMC therefore needs to look at two standards for telephone service: 1) the engineered blocking to its main switching complex; and, 2) the service standard with blocking, reneging, and time on-hold to each of the ten ACD appointment sites. The queue size, redialing, and reneging at the busy ACD sites on Monday morning can influence the variability in demand traffic estimates and confound attempts to size the trunk system for WHMC.

With this last point in mind, and since the present system at WHMC does not allow the pooling (or augmenting) of flexible servers, the literature points to increasing the efficiency of answering calls and allocating appointments only by adding servers to the busiest site at the particular time an additional server is available. However, a detailed description of telephone traffic flow over time by ACD site

has not been readily available until this study attempted to determine demand patterns. It was found that distinctive characteristic demand patterns existed for most of the ACD sites. Callers attempt to reach the appointment site in large numbers at the time when appointments are first available, and the large number of calls blocked (interflow) at the ACD queue suggest significant redialing. Callers know they must get in early to get an appointment or they will have to wait until the next week to try. They may tend to redial more frequently than the callers referred to in the literature, and so redialing may have a larger effect on demand. Regardless of the attempts to solve the telephone access problem in the past, the use of demand forecasting across the routine business system and the ACD sites to obtain a conglomerate view of characteristic demand, peakedness, and blocking was not methodically applied at WHMC as the literature would prescribe.

Besides not applying the demand data to forecast telephone traffic demand and its peaks, neither was it methodically applied to help determine alternatives for ACD queue length or total system trunk group requirements. An estimate for trunk group sizing from Southwestern Bell referenced only the system blocking on a Monday morning from eight to nine o'clock for a single week of observation data from Monday, 4 April to Friday, 8 April 1988. The data were collected to evaluate the impact of 30 trunk lines added to

the WHMC system the month before. The contractor's estimate was based only on point data versus the 20 single-hour measurements recommended in the literature. Furthermore, the contractor used the telephone industry standard of one percent blocking probability and the Erlang formulae to arrive at figure of 167 trunks required for a system blocking probability of 0.0094 percent. Clearly, this was a demonstration of over-provisioning trunk groups based on peakedness and variation in demand traffic; Monday traffic in that week was fundamentally different and significantly greater than the other days of that week (Figure 1 in Chapter II illustrated WHMC telephone traffic demand and blocking for the workweek of 4 to 8 April 1988). An attempt to balance the telephone traffic demand by moving some ACD arrival streams to other days of the week might allow a more consistent daily traffic load for the system which might require fewer trunk lines.

WHMC's not-for-profit nature as set in the framework of the military superstructure denudes it of the clear cost/revenue environment that many cases in the literature boast. Additionally, WHMC often cannot meet the level of health care demand requested (appointments fill-up before calls for them stop), but only provide as much care as possible on its share of the budget proffered by Congress. Currently, WHMC implicitly sets the service standard by only having one or two clerks attend the phones at the busiest

ACD sites even on the peak demand times. While industry seeks to minimize the cost of sustaining a service standard for customer demand, WHMC cannot afford to let customer needs dictate health care service levels. Hence, maximizing the efficiency with which it allocates appointments, and then defining how to deal with the rest of the callers which dial before the next queue of appointments opens (usually the following week) might be more appropriate for WHMC management. For instance, when all appointments are full, WHMC management could choose one of these two options: 1) clerks should fulfill the personal, yet monotonous, role of counselling callers who have missed an opportunity to get an appointment when all are full; or, 2) an answering machine could be used which provides the caller with a message stating "all appointments are full...(and some information and emergency numbers)", and the clerks' time could be used for other duties.

Commercial enterprises use ACD systems and computer databases to efficiently distribute a single queue of calls to numerous operators to conduct service, while WHMC uses the ACD system more as waiting room so that the majority of callers do not get busy signals while one or two clerks service the calls at a particular site either with a computer, a log book, or both. The method used to make an appointment depends on the state of the art of the hardware at the site and the predictability of the doctors' schedules

due to training, teaching, or other duties. It has been proven that the clerks using the computerized appointment system have a faster service rate for callers. They do not have to continuously search through a folder for the next open time slot, and their accounting tasks related to appointments are automatically computed, thus giving them more time to do other tasks at the ACD site. Further, the standardized process for using the computer to make appointments would allow clerks with similar understanding of medical terminology and procedure to be flexibly used between similar ACD sites.

Although WHMC would like to guarantee small waiting times on-hold and little renegeing, its ACD sites are not configured to afford such service standards. ACD queue lengths have been set to six under the assumption that at a constant service rate of two minutes a call no one will wait more than twelve minutes. Clerks at the decentralized sites are of limited supply due to manpower work standards and budget ceilings, and management at each ACD site routinely use the clerks for many other tasks as befits the responsibilities of a 906X0, medical administrative clerk. Unlike a commercial enterprise where ACD systems often have dedicated operators, problems in answering the telephone at WHMC arise when a site is dependent on the clerk for too many things at once (as on Monday mornings). This can be related to the literature on service times and interruptable

work, but the lack of a cost/revenue structure and the absolute cap on manpower make this situation unique. Getting the phone answered is even more of a problem when the clerk is on vacation or sick leave. Since it was observed that most of the sites share this phenomenon, the scheduling and staffing routines found in the literature might be useful if clerks can be cross-trained to provide a source for flexibility between a few sites with different peak demands. This flexibility would allow augmented manning of peak demand periods to the point that telephone and computer hardware and physical space at the appointment areas can provide. To reduce the load on the telephone system the queue lengths should be adjusted downward until observed renegeing goes to zero. This action would reduce the amount of caller waiting and free more lines for other traffic by not allowing calls which will not be answered (reneges due to excessive waiting) from congesting the system and increasing overflow.

Easy to implement scheduling and staffing rules for the ACD sites would be most useful. Complicated tables and graphs would be senseless for possible staffing levels of only one, two, and possibly three clerks against the magnitude of peakedness that present demand patterns indicate. A heuristic for scheduling stand-by servers could be devised either on a blocking, a renegeing, or a time-on-hold standard (or a combination thereof), so long as clerk

performance was ultimately judged with an idea of an upper limit on their productivity (e.g. 25-35 calls served per hour per clerk). The achievement of the goals of the heuristic could then become a feedback mechanism between the ACD data source in the communications branch and the ACD site supervisor and clerks in order to grade service and evaluate clerk performance by site. However, influencing the effect on caller perception of not getting the appointment for which they called is beyond of the scope of this study, and it may well be beyond the public relations capabilities of the military community.

Decentralization is inherent to the diverse nature of the different clinical specialties at the various ACD sites at WHMC, so consolidating to a central appointment system seems out of the realm of possible solutions. A July 1988 interview with the civilian supervisor of the central appointments office at Brooke Army Medical Hospital, also in San Antonio, Texas, revealed that it is possible for a fulltime civilian cadre (without medical training backgrounds) to staff the appointment lines for the clinics at that hospital with the aid of a master scheduler, a fully computerized appointment system, and reference books containing medical questions with which prospective patients are screened to be appointed to a specific clinic and doctor. However, it was determined that these personnel accounted for roughly only one third of all the appointments

made, and the clerks at the clinics filled the remainder and tended their other duties. The savings in manpower are questionable.

Findings of the Literature Review and Discussion

Several significant findings from the literature and discussion have influenced this study of WHMC's phone system. Two problems have been identified: 1) overflow of the entire system when all new traffic is blocked; and, 2) blocking at the ACD sites when they reach capacity, interflow. The importance of maintaining a database of demand traffic for the purposes of forecasting and as feedback to operators and managers is evident. The case for studying the busiest hour of the month has been made. The problems of variation in the data for establishing true demand and of determining the appropriate sample size for estimating demand traffic have been noted. WHMC has been found to be unique in the constraints it has on its phone system and on its resources. Simulation has been chosen as the best means of approaching the problem after careful examination of many methodologies. Most of all, the measure of service at WHMC and the ACD sites has been reduced to the throughput of callers. Caller perception was found to not depend as much on how long they waited in queue, but how their time was occupied while waiting. Efforts in this

research were made to maximize the number of ACD callers which quickly get either an appointment, or notification that an appointment is not available, while reducing the amount of overflow to the telephone system.

IV. Data Reduction

Chapter Synopsis

The hourly data for all the ACD sites for several months were collected, and the parameters for the simulation model were derived after much data manipulation. There were several peculiarities in the data; some were due to the way in which the ACD software collected the data, while others were due to clerk behavior at the individual appointment sites. It was observed that there was a distinct monthly, weekly, and daily demand pattern for most ACD sites, and that the peaks in these patterns were highly correlated to the times when appointments were available to be booked at the individual sites. A new measure of demand and accessibility was defined which reflected the importance of the interflow calls at each ACD site. It was found that the monthly pattern was very similar from month-to-month in both the shape and amplitude of the traffic of calls accepted into the queues at each ACD site. This last finding led to establishing the cycle of the ACD site demand traffic to the telephone system as monthly. The busiest time of the monthly pattern was observed to be 8 to 9 A.M. on Monday morning of the first full workweek of the month. The non-ACD traffic for the system was estimated from a study of the

downtown switching complex by Southwestern Bell, which covered the first full week of April 1988.

Definition of Terms

Below are a few terms important to the description of the data reduction effort. These terms relate directly to the ACD hourly reports from which the data for this research was gathered. A sample ACD hourly report is in Appendix A.

Peg-count Data: Call Intensity and Congestion. Simple observation data collected by the ACD system software from the SL-1 digital switching complex of WHMC was reported on the hourly summaries as a single number for each site. These numbers represent the intensity with which callers were dialing to reach an appointment clerk, and how their attempts succeeded. When call intensity was very high to any site, or to WHMC in general, system congestion due to call intensity could be present. Definitions of the most relevant of the frequency data used to describe telephone traffic at WHMC are below.

- 1) Calls Accepted: The number of calls that were accepted into the queue at an ACD site. These calls were allowed to wait (were delayed) in the queue for a server.
- 2) Calls Answered: The number of calls which actually reached the appointment clerk.
- 3) Calls Abandoned: The number of calls which were accepted into the queue, but which hung-up before reaching an appointment clerk. These were also referred to as reneges from the queue and as a loss to the system.

4) Interflow: Calls which arrived at the ACD site when the queue was full (six calls waiting and however many calls then being served) which were routed to the "all circuits busy message." This was virtually call blocking, but it occurred at the individual ACD sites, and was a loss to the system.

5) Overflow: Any call blocked from WHMC at the downtown switch, whether its destination was an ACD site or a non-ACD phone, was labeled as blocked to the system with no distinction for destination available.

6) Number of Attempts: The frequency count of calls arriving at the downtown switch with a destination into WHMC. Separate from calls blocked, this was the total number of calls which actually made it to the SL-1 digital switch inside WHMC, and was considered the "total demand traffic for the WHMC phone system" for the time period covered.

7) Non-ACD Traffic: All telephone traffic into WHMC not destined for an ACD appointment site.

Except for calls which were on-hold across the hour mark of the digital clock, calls accepted is equal to the sum of calls answered and calls abandoned for each hour of the ACD hourly reports. Calls accepted plus the interflow for each ACD site represent the total demand traffic for an individual ACD site in the hour of measurement. However, calls which never reached the ACD software were also important to describing the demand traffic for the system. A look at these types of calls came from the traffic study conducted at the downtown switch for the first full week of April with reports generated on the half hour, 24 hours a day. Logic dictated that when one specific hour of the first week of April 1988 was considered, the demand traffic for non-ACD sites would be the difference between the total

demand for the WHMC phone system and the sum of the demand traffic at the individual ACD sites.

Time Spent Data: Time Intensity and Congestion.

Information on the duration of calls was collected for both ACD and non-ACD traffic. Aggregate talk time was measured for the WHMC phone system on the half hour in the downtown study. Calculation of the non-ACD talk time follows the logic above for non-ACD call traffic. Talk time intensity refers to the duration of the average call for a given period of time. Time congestion refers to the phenomenon that as callers stay longer on the phone (talking or holding), fewer lines are available to service new calls. As this congestion grows, more blocking to the system and at each ACD site can occur.

1) Average Speed of Answer (ASA): The sum of time spent in the queue by all callers who reached a clerk divided by the number of these callers.

2) Average Time Abandon: The sum of time spent in the queue until the caller hung-up before reaching a clerk (reneging) divided by the number of these callers.

3) Direct Call Processing (DCP): The average time the clerk spent on the phone with a customer.

4) Post Call Processing (PCP): The average time the clerk set the phone station to "not ready" for the purpose of completing paper/computer work to make each appointment. The "not ready" setting disables the ACD software from forcing the next call to be answered by the clerk, and allows the clerk to finish processing one caller's request before answering the next.

5) Non-ACD Talk Time: The sum of call times for all telephone connections made through the downtown switch of any measurable duration (actual conversations, phone rings but not answered, etc.) made to anywhere in WHMC other than an ACD site.

Problems with the Data. There were three main sources of difficulty in reducing the ACD data to the parameters for the simulation model. These sources were the ACD statistics and software, the behavior of the clerks, and the behavior of the callers. All of these sources inherently interacted and lent variability to the data, sometimes with a degree of magnitude of difference in a statistic observed for a common hour between similar days. These swings in the data values were indicative of the real situation observed during the interview process, and were considered when applying the results of the simulation model to the WHMC phone system.

ACD Data

There were three reports generated by the ACD software on an hourly basis (Appendix A) with a daily totals report being printed at the end of each day. Each report was indexed by the telephone extension of the individual ACD sites with an ACD system total at the bottom of the report. While there were many fields of data in these reports, only the data elements deemed to pertain to this study are of interest here. The first hourly report gave these statistics: 1) the average number of clerks working; 2)

the number of calls answered at the site; 3) the average speed of answer (ASA); 4) the direct call processing time (DCP); and, 5) the post call processing time (PCP). The second hourly report provided these: 1) the number of calls accepted; 2) the number of calls which abandoned; 3) the average time to abandon; and, 4) the amount of interflow at the site. The last hourly report, described the number of calls answered and average time spent data for the individual phone positions at each ACD site from which the time averages for the first two reports were calculated. The "daily totals" report was the average for the day of the first two reports.

The ACD Software: Accountability. The ACD software made assumptions about how its hardware was used when it compiled the hourly ACD site reports. The software assumed that a server was dedicated solely to answer the phone without server vacations or interruptions. When the clerk needed to be away from the phone or was interrupted for whatever reason, it was most common for them to put the phone-set in the "not ready" mode. This was consistent with the ACD software idea of post call processing, but when vacations or interruptions became long, the software still would show the position as being staffed and would credit the post call processing statistic with the work. The most misleading effect was that when the hardware was set to "not ready", when in fact the server was away the entire hour,

the statistics for the number of servers, and server utilization were inflated. This also gave the impression that the servers were extremely slow at their duties, since two agents could be shown as working the hour when the number of calls actually served would be a single clerk's effort.

On the other hand, the phone-set could also be put in the "make set busy" mode. The statistical effects of this were to remove the phone-set from registering in the software. The "not ready" mode assumed a clerk was present, but temporarily indisposed, while the "make set busy" mode shuts down the position assuming an extended vacation. If all positions at a site were set "not ready" callers would queue, but would have no one to answer; thus, experiencing long waits. If all positions were set to "make set busy", the hourly report showed asterisks for the site, and calls were diverted to the night message as if the site were closed for the day. In both cases the statistics lost value for direct interpretation and for creating model parameters. Further, both settings can cause callers distress, one by long waits, the other by the perception that the site was officially closed during the business day.

The ACD Software: Averaging. The most severe limitation inherent in the ACD data for constructing the simulation model was that only hourly averages were reported for the time spent statistics, and only point observations

were available for the peg-count statistics. This limitation meant that only assumptions could be used for determining the functional form of any of the arrival or service distributions. The downtown switch data was reported on the half hour and revealed a higher intensity of traffic from 0800 to 0830 than from 0830 to 0900, so the hourly average of the ACD data misses the actual distribution of the traffic. The large differences in the averages for arrivals across the hours of the busiest morning for any one site indicated a nonstationary arrival process, but the only way to assess how the calls arrived within hours across the day was by interviews with clerks.

The clerks agreed that when busy, the queue was constantly occupied, and that when slow, calls came more in spurts of activity rather than uniformly over the hour. The clerks further agreed that the length of a conversation was extremely variable being low for a routine appointment for a caller prepared with necessary appointment information to very long when trying to calm and help, or explain things to, an angry caller. When similar hours of comparable data were analyzed, there was a great deal of variation in the point observations. This analysis led to the argument for using a poisson arrival process and exponential service and post call processing times made earlier as being a simple and conservative approximation for the simulation.

ACD Software: Interpretation. Another aspect of the averages reported by the ACD software had important managerial significance. Off-site managers trying to interpret these ACD averages over this environment for short periods of time could easily draw faulty conclusions on clerk performance unless sensitized to the nature of the ACD statistical gathering routine. For example, let only fifteen calls arrive in a slow hour, but allow them to be grouped in a small amount of time. Furthermore, let a single clerk answer five of these calls in succession in ten minutes of work, while the other callers renege due to waiting in queue while others were served. An uninformed manager might complain that the clerk only had fifteen calls in the entire hour, but was so lazy only thirty percent were answered, and even those five had a long wait in queue.

In reality, the clerk was very busy when the calls arrived, working non-stop until no one was holding in the queue. Although the renegeing behavior would tend to lower time in the queue for callers answered, assume the initial five callers stayed in line and the clerk took exactly two minutes to service a call. The ACD statistic for the average speed of answer ($0+2+4+6+8/5=4$ minutes average wait) does not describe how callers were actually treated. Three were spoken to within four minutes, while two had longer waits, but only one caller realized the average. The averaging of wait times in the ACD hourly reports tends to

smooth out the underlying distribution of experienced wait. Also, it covers up the worst case of experienced waiting from management while the clerk gets to talk to someone who has been holding for a long time (and then might not get the help/appointment for which they called). The situation of a few calls arriving in spurts for the hour clearly can be seen differently from the clerk's and an uninvolved manager's point of view. The numbers involved could make the clerk look even worse in a busy hour, even if the clerk worked 99 percent of the time, the percentage of calls answered to calls accepted and interflow might be very low.

Clerk Behavior and Supervision. Clerk behavior and their site managers had a large role in making it hard to reduce the ACD hourly report data. Specifically, the lack of scheduled service, a policy allowing interruptions, low clerk dedication, and clerk misuse of the ACD hardware added variation to the data. The lack of scheduling resulted in some sites opening at different times and even not staffing the phones at various times of the morning. In addition, across the five comparable Mondays of data, there were often different numbers of clerks working across the hours of the day for most sites. Management policy allowing interruptions led to fewer calls being answered depending on the amount of time lost from answering the phone. Clerks who wanted to shirk their duties could easily just not answer the phone when not watched by supervisors, and they

could also seek unnecessary interruptions to get away from the phone answering task. This disrupts data from sites with only one server while masking performance indicators of sites with two servers. Finally, the clerk can skew the statistics collected by the ACD software by misusing the "not ready" and "make set busy" features of the ACD hardware. During the interviews, many clerks acknowledged that many types of business interruptions, as well as their own desire for a break, routinely led them to set their ACD positions to "not ready" or "make set busy" for various intervals of time throughout the day.

Caller Behavior. The way in which calls arrived, held-on, and reneged were also sources of error which added variability to the data. The downtown study described call attempts to the WHMC system in half hour increments, and there was a marked jump in call attempts on Monday 4 April 1988 between 0700-0730 and 0730-0800 (404 to 1738 attempts). This jump was attributed in part to callers attempting to get into the ACD appointment sites which were booking appointments and opened at 0730. The jump in activity lead to the conclusion that callers were aware of the opening times of the ACD appointment centers, and that they make their attempts accordingly. With regards to a simulation model, this meant that there was no real transition period; just before 0730 things were quiet, and then activity leapt to a high intensity.

Other types of caller behavior had an affect on the arrival data variability. It was speculative to guess at how many callers tried to game the system by calling just before the site was scheduled to open in order to be the first one in the queue. It was also a guess as to how many of the total call attempts to the ACD sites were repeated attempts. Repeated attempts could be made by callers who either received the "all circuits busy message", received the night message when trying to game the opening of the ACD site, or reneged from the queue for whatever reason and were trying to get in again. As discussed in the literature review, callers tend to repeat their calls until they feel the probability of being successful was too low to bother to continue. A survey of callers at the Primary Care site on three different Monday mornings asked callers the question, "How many times have you called today for this appointment?". The survey revealed that several callers stated they had called as many as ten to fifteen times, and that less than fifteen percent of the callers polled during the busiest hour reported getting through on their first attempt. Repeated calls tend to drive the number of calls accepted, and interflow higher; thus, making it hard to establish the actual telephone traffic demand at the sites. Reneging had a similar effect on the number of calls abandoned, but reneging also could affect the time spent statistics.

The decision to renege or hold-on was the last in the series of caller behaviors to affect the data. The majority of callers seemed to understand how congested the WHMC system and ACD site queues could get on Monday morning as well as how appointments could be given out in a short time of a site's opening. With this knowledge, many callers when successful in getting through to an appointment site would stay on-hold until served. At 0730 when some queues open, there are often already six calls in them. These callers hold as long as it takes or renege. With this behavior pattern, it was no wonder that wait times of forty minutes were common before the ACD system went from no cap on the queue length to allowing only six to queue at any ACD site. What a caller did after gaining entrance to the system impacted both the peg-count and time spent statistics. Callers who held-up telephone line resources, but eventually reneged loaded the system for no reason.

ACD and WHMC Hardware. The two greatest limitations to the scope of this study were caused by the hardware. First, the data from the ACD software was not stored electronically, but was only available from daily paper printouts. The effort to gain a large database of the different clinics for different days of the week and across different months was very cumbersome and time consuming. The second hardware constraint was the addition of the thirty commercial lines in March 1988. Though this may have

helped to alleviate the system congestion somewhat, it did not cure it. The data collected up to this time lost comparability to data collected after this date, and data for the system had to start from zero and build up over the following months to get the five point observations used for this study (April to August 1988).

The Downtown Switch Study Data. The downtown data came from the first full week of the month of April 1988. From Monday to Friday at half hour intervals, data on the number of attempts, aggregate call duration, and overflow were collected. The only overflow for the entire week occurred on Monday morning from 0730 to 1000 when the system was most congested (Figure 3.). This finding along with findings from the literature review served as the basis for targeting Monday morning from 8 to 9 A.M. as the time when a change for the better would best be appreciated and best aid accessibility. The half hourly reports revealed that traffic from 0800 to 0830 was more intense than from 0830 to 0900, but ACD data is only output on the hour. Little advantage can be made of this knowledge about the actual demand distribution in the simulation model, since ACD data is reported by the hour.

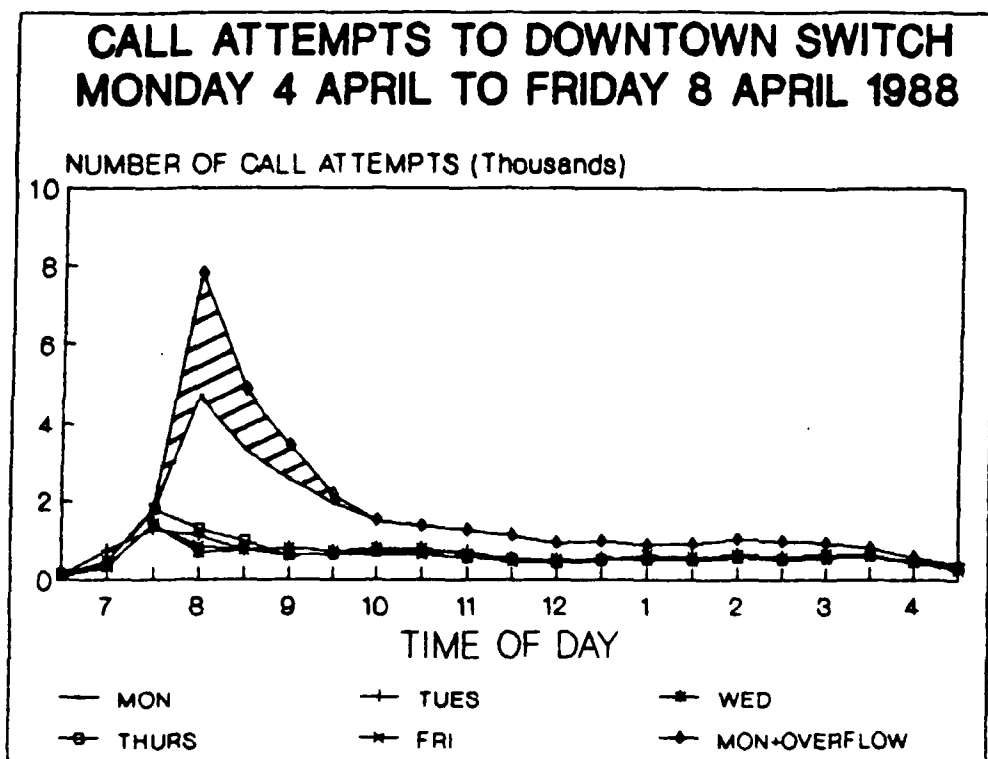


Figure 3. Weekly Call Attempt Pattern

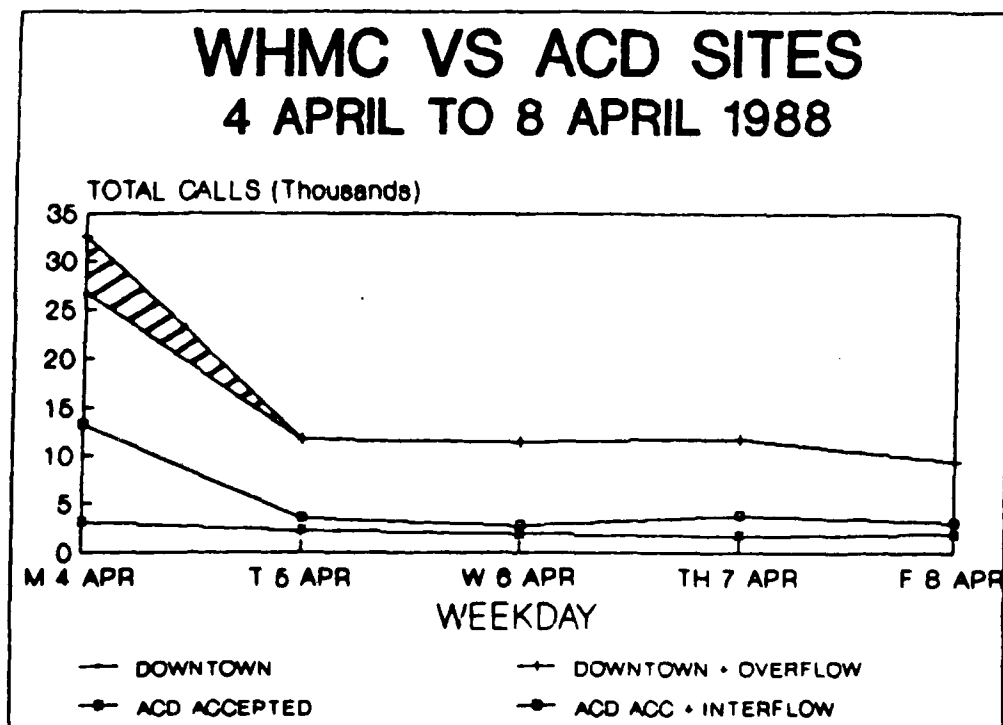


Figure 4. Full System Data

Demand Patterns

The only overflow traffic (shaded region of Figure 3) occurs on Monday morning between 0730 and 1000 hours. The full system data in Figure 4 shows that the ACD sites' traffic is a large part of this week's activity to WHMC.

While the number of calls accepted to ACD sites looks steady over the week, the interflow pattern for the ACD sites is peaked and correlates to the pattern for the total traffic for WHMC going through the downtown switch. The interflow pattern also shows that there is a great deal of day-to-day load variation in offered load to the system. As will be shown below, ACD site traffic peaks follow the time of week or month that appointments are made available. The above observations led to the hypothesis that if ACD demand traffic patterns could be predicted and controlled by changing the time appointments are made available, then system traffic could be altered to lighten traffic conditions on Monday morning by changing Monday morning ACD traffic patterns.

Only the one week of data was available for the downtown switch due to the expense of collecting it. However, much data for the ACD sites, as described earlier, spanned the period of March 1987 to September 1988. Data were stored as an aggregate statistic for each site's daily and weekly calls accepted totals, while interflow statistics

were not routinely recorded until the period beginning 25 April 1988. While it is obvious that calls accepted to the ACD site can be much smaller than the true amount of calls (calls accepted plus interflow, and the unmeasurable number of overflow calls destined for a site), any pattern across the months can still be a valuable indicator of traffic. Some data for entire days and even weeks were missing due to printer problems with the ACD hardware, but it was easy to find several months with few missing days of data and interpolate to fill missing data in order to examine monthly patterns. Data missing due to legal holidays were left as zero values.

Inter-month Data. Two trends in the data were necessary to add credibility to the choice of Monday of the first full workweek of the month as the busiest day for ACD traffic and for WHMC: 1) the calls accepted for the first Monday of the first full workweek of the month had to have more traffic than other days of the month; and, 2) there had to be an observable trend of calls accepted being patterned around the day appointments were available at the different sites. Figure 5 shows the calls accepted data summed across all ten ACD sites for six months (March and July 1987, and February, March, May, and July 1988). The data represent four full workweeks in each month. It is easy to see that Monday of each week is the busiest day, and that the first Monday does have more activity than the others in the month.

Additionally, the weekly ACD traffic pattern seems to start with the highest demand on Monday with a steady decline to Thursday and then an increase of traffic again on Friday. Note that traditional Monday traffic shifts to Tuesday when Monday is a legal holiday. Figures 6 to 9 are from the data used in Figure 5, but now the graphs are of individual ACD sites instead of being summed. These sites were chosen due to the high number of daily calls accepted and because they have different policies for opening appointment calendars.

There is a distinct observable relationship between ACD calls accepted and the time the ACD site makes appointments available. Figure 6 shows peaking on Monday when the Primary Care (PCARE) site opens its calendar each Monday for the week two weeks away. Figure 7 shows the Tuesday peak of General Surgery (GSURG) which books on that day for the week six weeks hence. Figure 8 is the high first week demand pattern of Gynecology (GYN) which opens all of the nearly 300 next month's Pap smear appointments on the first Monday of the first full workweek of the month. The peak on the second Monday of the GYN graph shows the effect of 4 July 1988, which was a Monday. Figure 9 dramatically illustrates that the Dermatology (DERM) clinic opens its appointments on each Friday. The other ACD sites had similar demand patterns for the days they booked appointments with the exception of the Obstetrics (OB) site which had no distinct pattern.

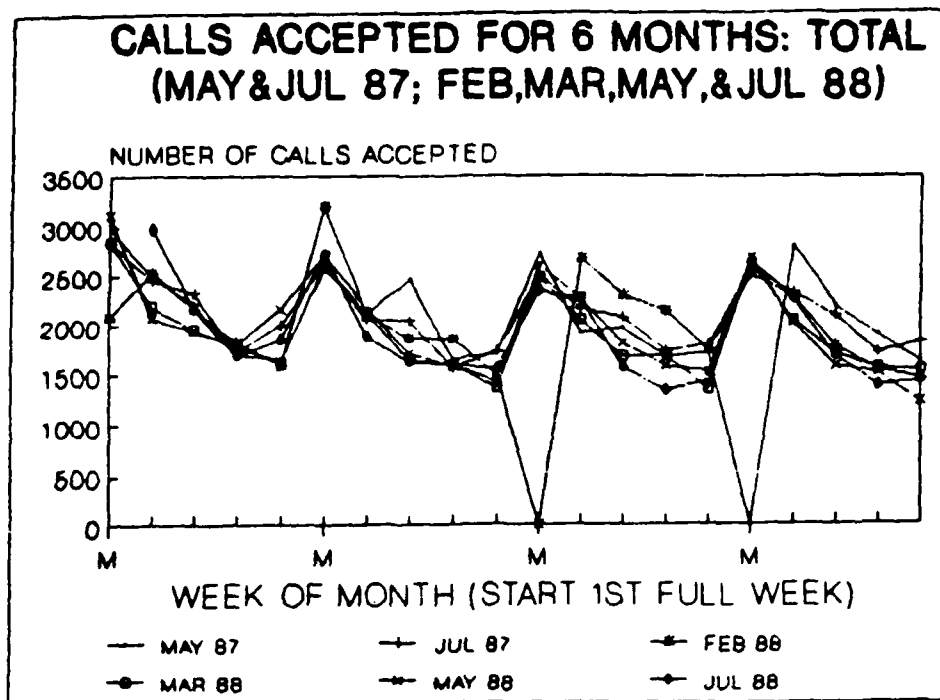


Figure 5. Long-term Demand for TOTAL ACD

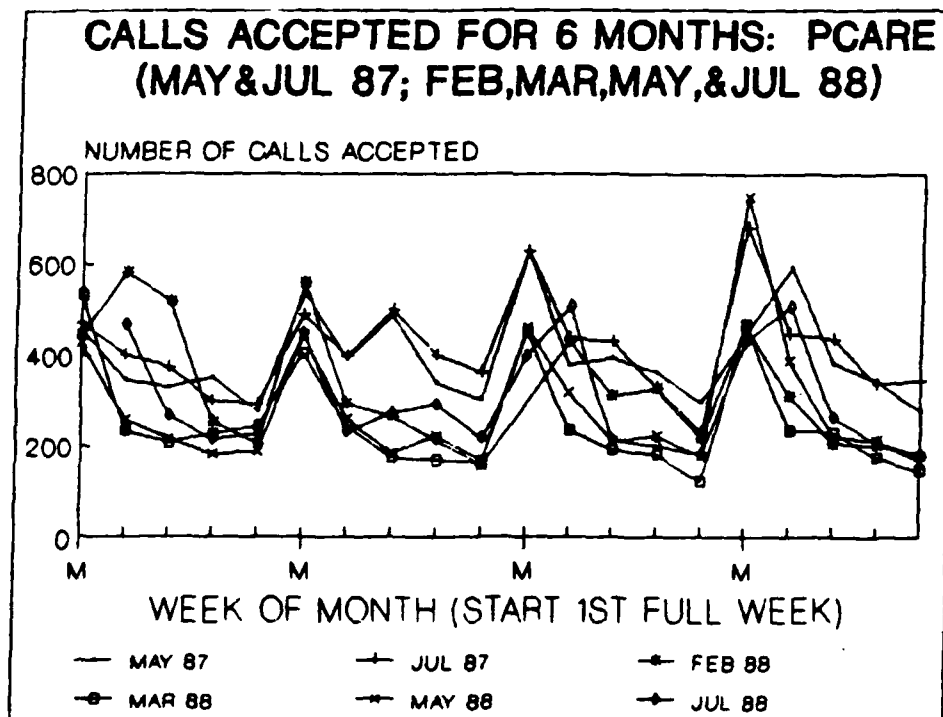


Figure 6. Long-term Demand at PCARE

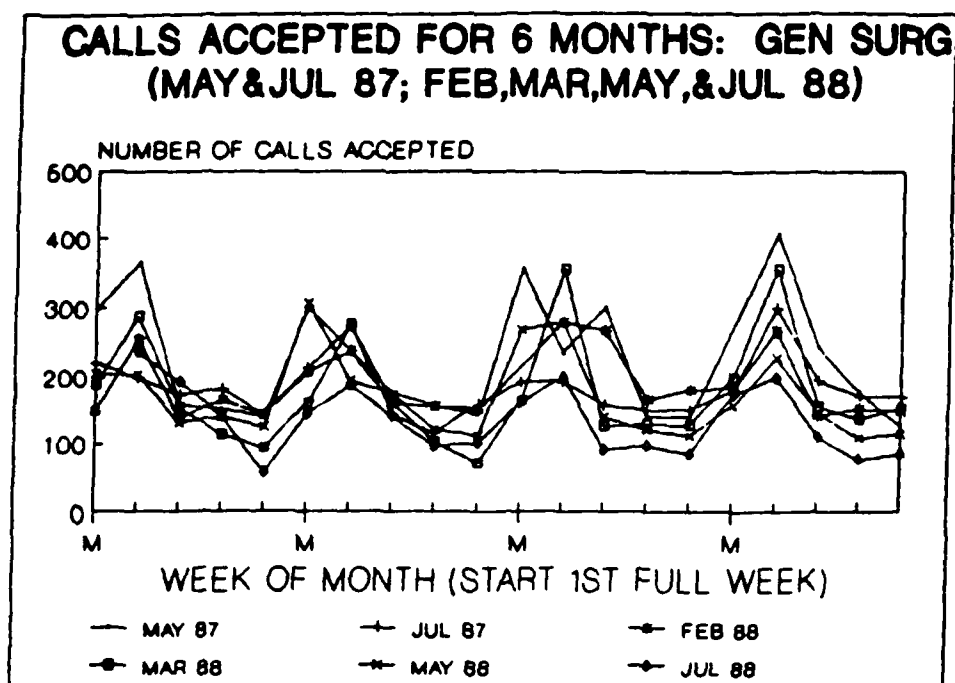


Figure 7. Long-term Demand at GSURG

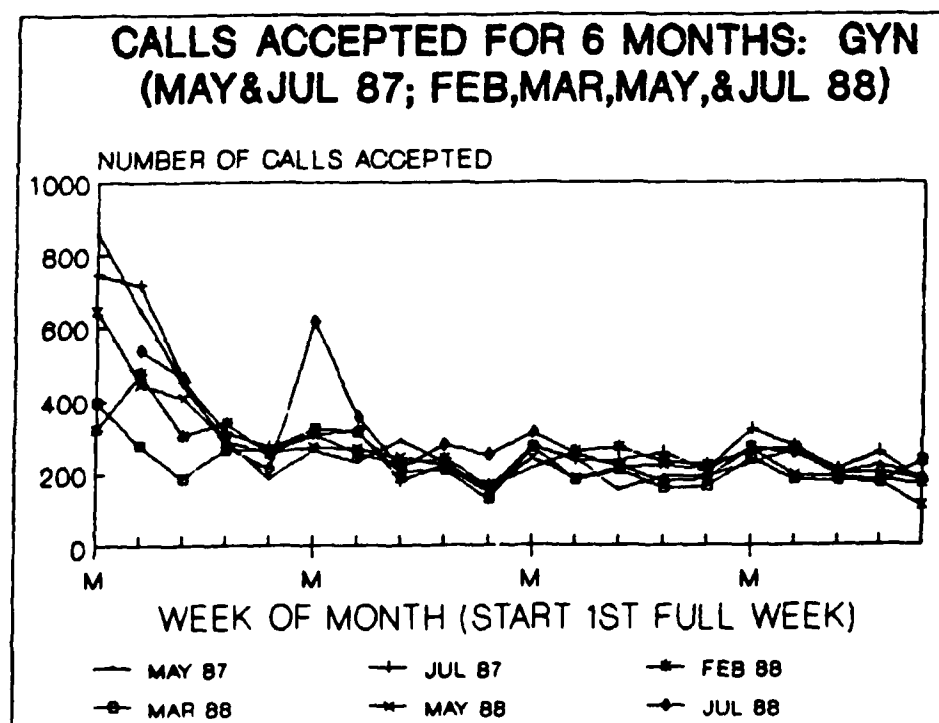


Figure 8. Long-term Demand at GYN

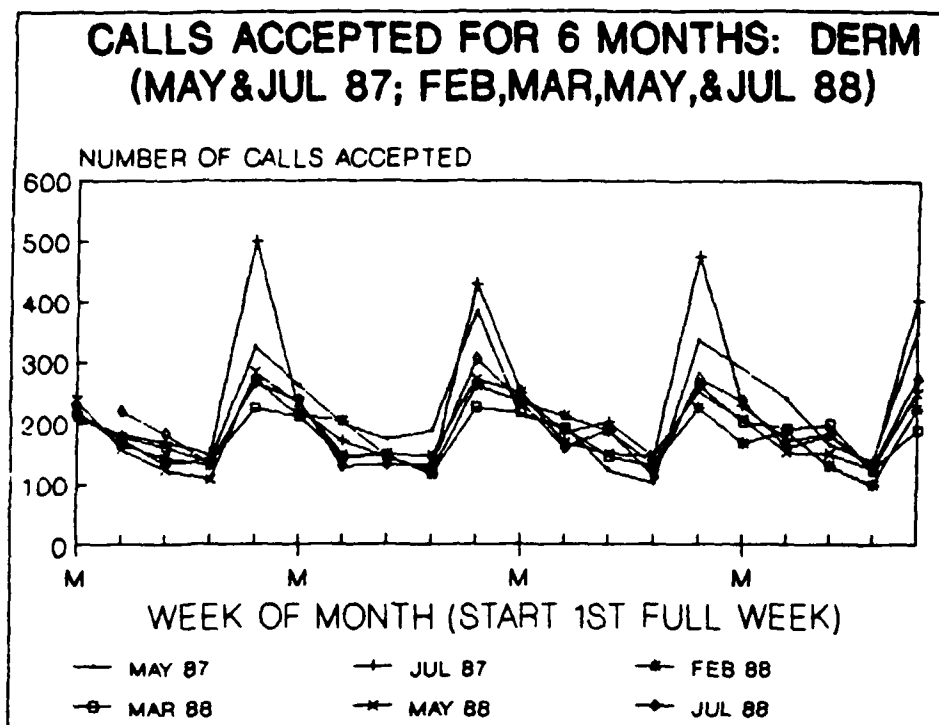


Figure 9. Long-term Demand at DERM

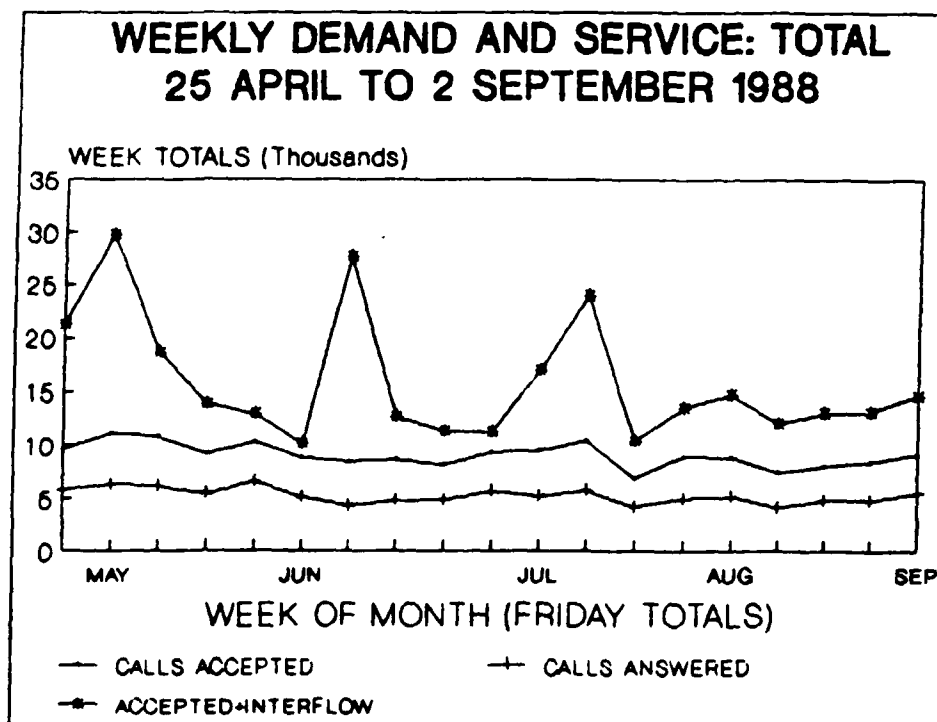


Figure 10. Demand Over Time for TOTAL ACD

Comparison between months allows checking for seasonality of demand by time of year, and also allows checking for the effects of the November 1987 change in queue length policy and the March 1988 addition of 30 commercial trunks. As mentioned above, calls accepted data does not fully describe demand, but there seems to be no seasonal variation in the months presented, nor is there variation between the years when like months are compared. December is said to be a slow month, but the data for that month are missing. The months of data cover the period before, during, and after the two changes to the WHMC telephone system configuration. Referring back to Figures 5 through 9 no apparent difference in the calls accepted statistic can be observed from either the change in the queue length at the ACD sites or the addition of 30 commercial lines to the system. This means that the ACD appointment centers have been operating at the upper limits of their current efficiencies in all the months observed.

Though the changes to the WHMC telephone system had no effect on the calls accepted statistic they did have visible results. Limiting the queue length definitely helped to decrease the time-on-hold of callers as evidenced by complaints of high long-distance phone bills dropping close to zero shortly after the policy was implemented. Limiting the queue length also kept the ACD calls from loading the system to the extent they had before. After the 30 lines

were added, WHMC personnel had noticed it was easier to receive routine calls. The effect of these system changes on caller perceived accessibility is speculative since true demand is unknown and since actually getting appointments also is an unknown factor. However, shortening the queue would likely increase the interflow statistic since fewer people were allowed to hold a shorter time, and adding lines likely lessened the amount of overflow traffic, so that more calls got through to their destinations.

Monthly and Weekly Patterns. To better view the actual demand and visualize the telephone's role in access to the ACD sites the interflow traffic statistic must be observed. Figure 10 shows the sums of all ten ACD sites' weekly aggregated statistics for calls accepted, calls answered, and interflow beginning 25 April 1988. These data were automated by personnel in the communications branch at WHMC, and were proof of their improvement in maintaining an active database on WHMC telephone traffic. The bottom line of the graph is the number of calls answered, but again this is no guarantee of getting an appointment. The distance between this and the next higher line, calls accepted, is the number of callers who reneged, or for some reason were in the queue at the ACD site but did not reach a clerk. Whether the wait was too long for these callers, or they were somehow disconnected, these calls do not appear to add much in favor of the perception of accessibility. The difference between

calls accepted and the highest line of the graph is the interflow calls. These are the callers who received the "all circuits busy" message, and are without a doubt being denied access to the appointment center they desire. The peakedness of the interflow area is an indication of the redialing phenomenon, where the peg-count data increases for each call attempt but overestimates true demand, since a single person trying without success three times would appear as three separate callers. Calls blocked at the downtown switch are all denied access, but it is impossible to determine their destination.

Figure 10 also illustrates a dimension of caller behavior. The weekly sum of calls answered and calls accepted seem stable, while the interflow statistic shows more traffic on the first week of the month and less traffic the remainder of the month. This pattern appeared earlier in the graphs of calls accepted, but now a more realistic view of access is presented. By looking at the data for the first week of May in Figure 10, it can be seen that typically less than 60 percent of the calls accepted into the ACD system were answered, and that only about 20 percent of all the calls destined for ACD sites were successful in reaching an appointment clerk (again getting an appointment is an unknown). However, looking at the data for the first week in August in Figure 10, it is easy to notice that the characteristic peak has been erased. The explanation is

simple, the GYN site changed its policy for booking Pap smears to making half on the first Monday of the first full workweek of the month and half on the following Monday (52). This is a clear demonstration of implementing a change in appointment booking policy and having a marked impact on the telephone traffic into the ACD system. A look at a particular month day-by-day is needed to show how traffic can vary over the workweeks in a month.

Weekly Patterns. The five week period from 4 July to 5 August was selected to portray the monthly and weekly seasonality of the ACD system data. Figure 11 graphs total ACD demand and service patterns for this period. The peak on the second Monday is the effect of the definition of first full workweek, since the month began with a holiday. Wednesdays and Thursdays are typically the slow days for the overall ACD system. Most ACD sites have spikes in their interflow pattern which coincide with the day they open the books for new appointments. Callers understand and behave consistently with the lottery nature of the first-come-first-serve appointment by telephone system at WHMC. Call intensity is high when appointments are first available and the intensity quickly decreases until the next time an appointment can be made. This pattern is evident regardless of the day of week (where a pattern exists). To use caller behavior to advantage in alleviating the Monday morning congestion at WHMC, the four ACD sites which book on Monday

(Primary Care: PCARE, Orthopedics: ORTHP, Gynecology: GYN, and Pediatrics: PEDS) are scrutinized for 4 April 1988, the day of available data for the downtown switch. Moving their combined demand peaks off of Monday is evaluated for the effect on the Monday morning WHMC phone system congestion in the simulation model in Chapter VI.

Busy Day Data. The demand and service patterns for 4 April 1988 for the ACD sites just mentioned tell different stories of accessibility. The interflow pattern is the best measure of demand, and some part of any overall evaluation of accessibility should include the interflow statistic, though the effect of redialing on true demand must be considered. Figure 12 is the 4 April activity at PCARE, and it shows a large peak in interflow from 7 to 8 A.M. (actually, PCARE opens at 0730). Figure 13 shows the morning interflow pattern for ORTHP for 4 April, while figure 14 shows GYN opening at 0800 and being flooded with calls the entire day with a secondary peak of calls arriving between 2 and 3 P.M.

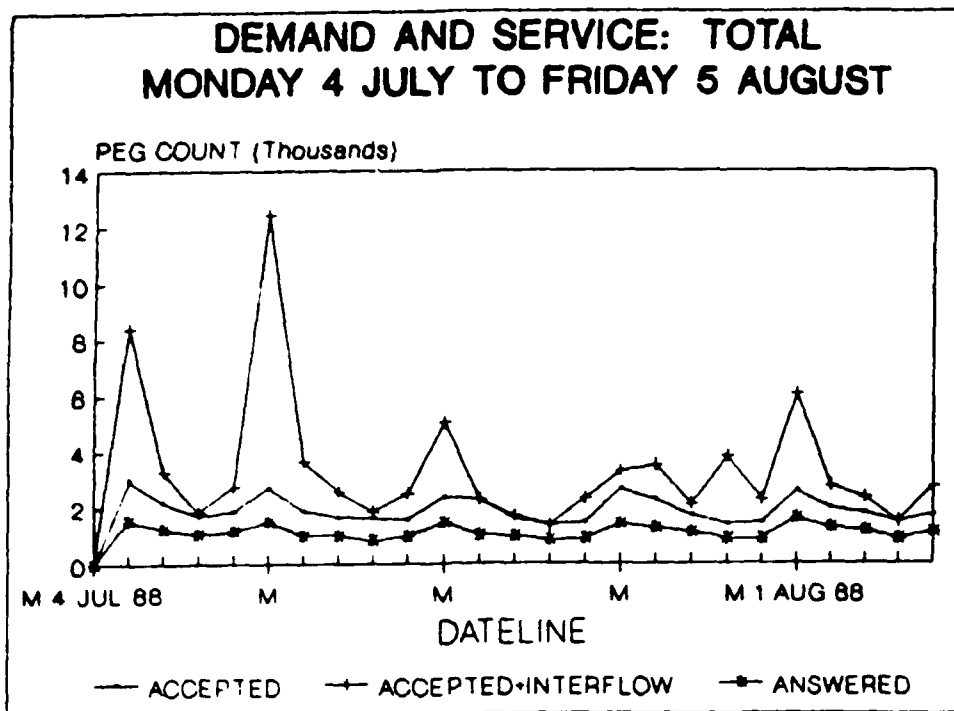


Figure 11. Monthly Pattern TOTAL ACD

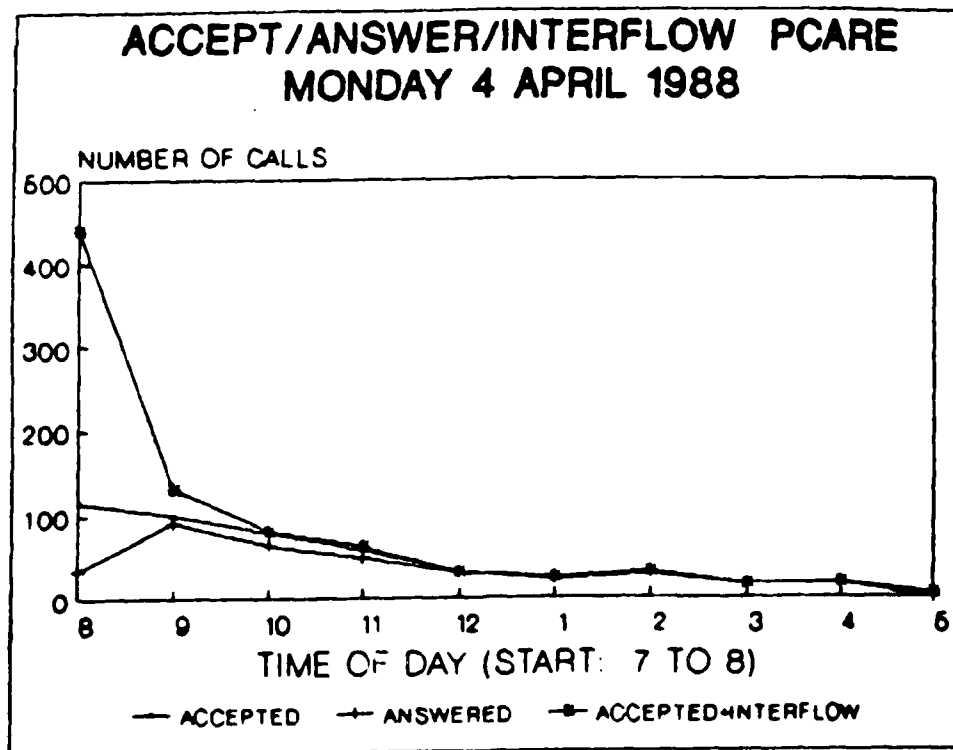


Figure 12. Daily Pattern PCARE

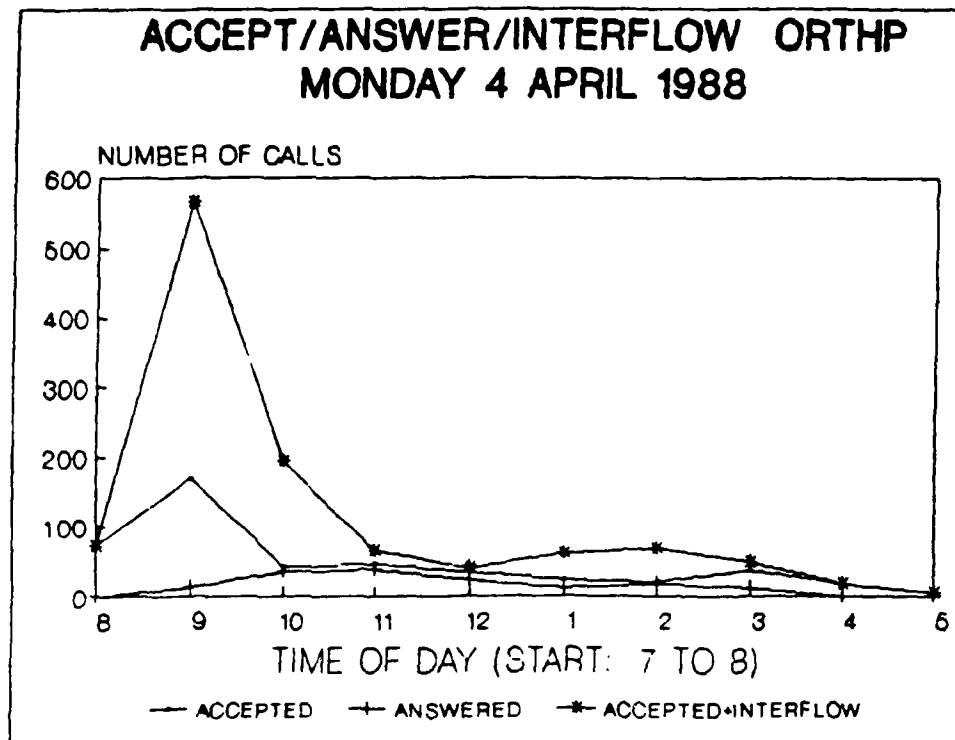


Figure 13. Daily Pattern ORTHP

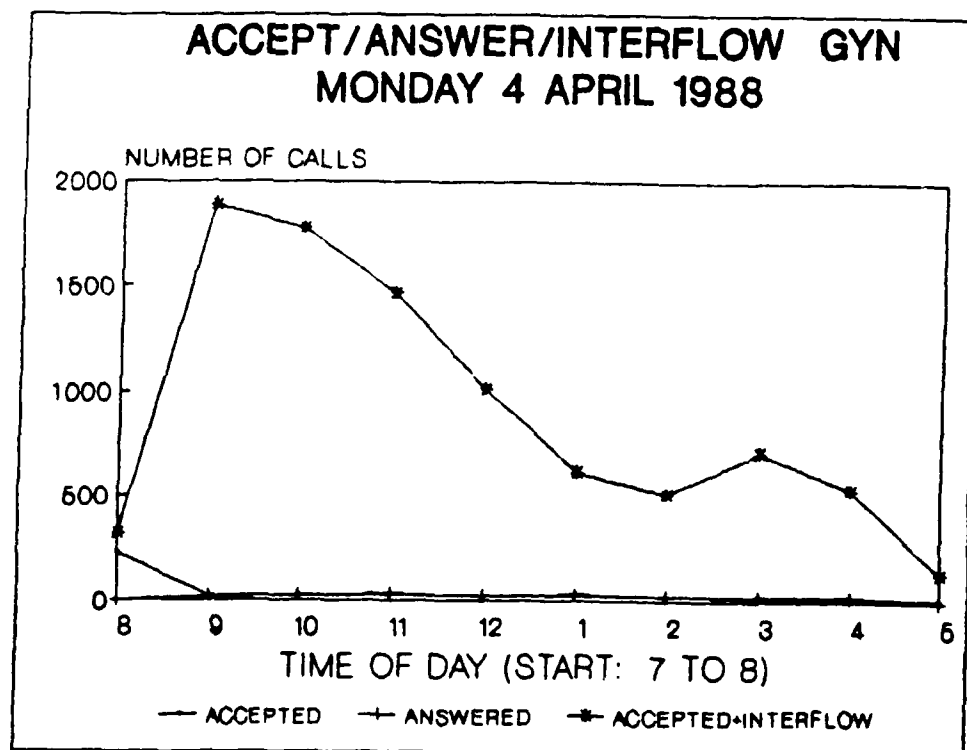


Figure 14. Daily Pattern GYN

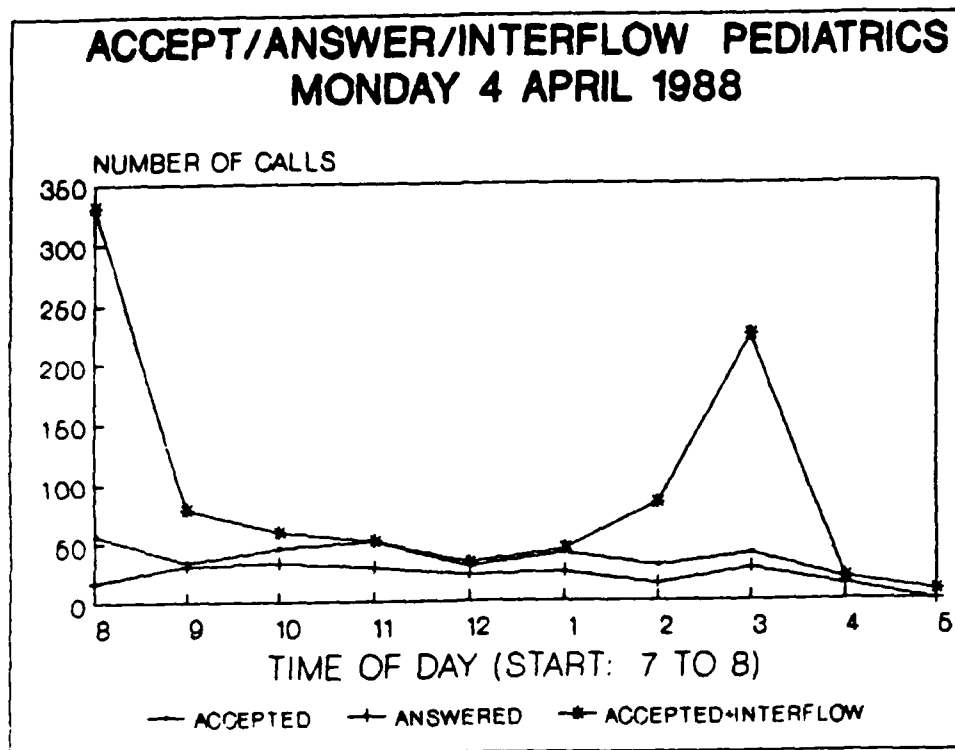


Figure 15. Daily Pattern PEDS

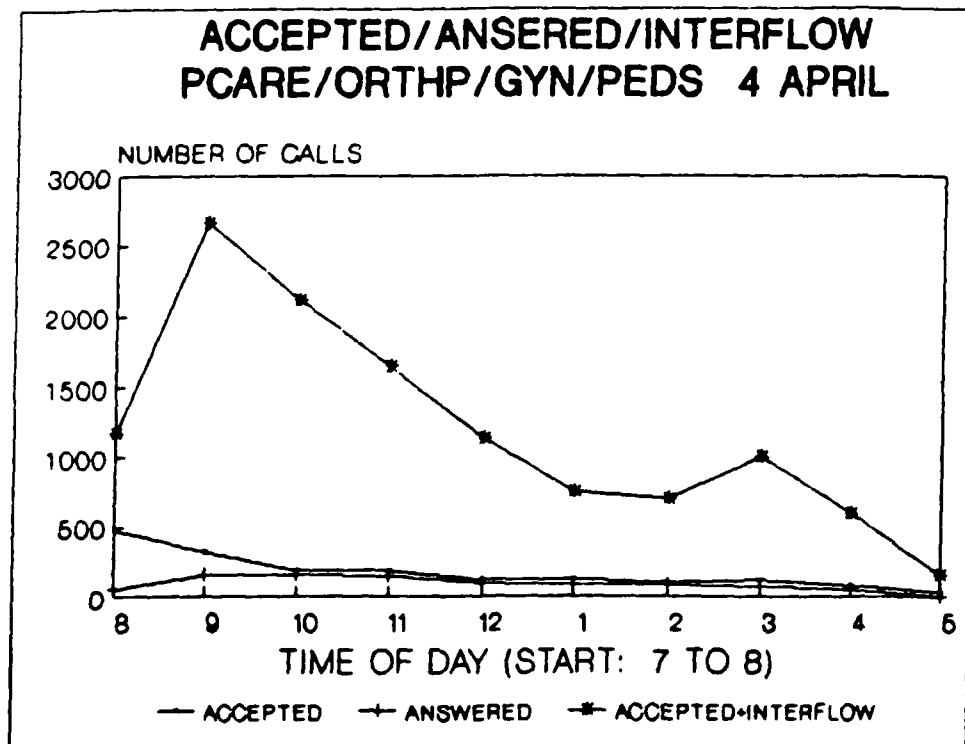


Figure 16. Monday ACD Traffic

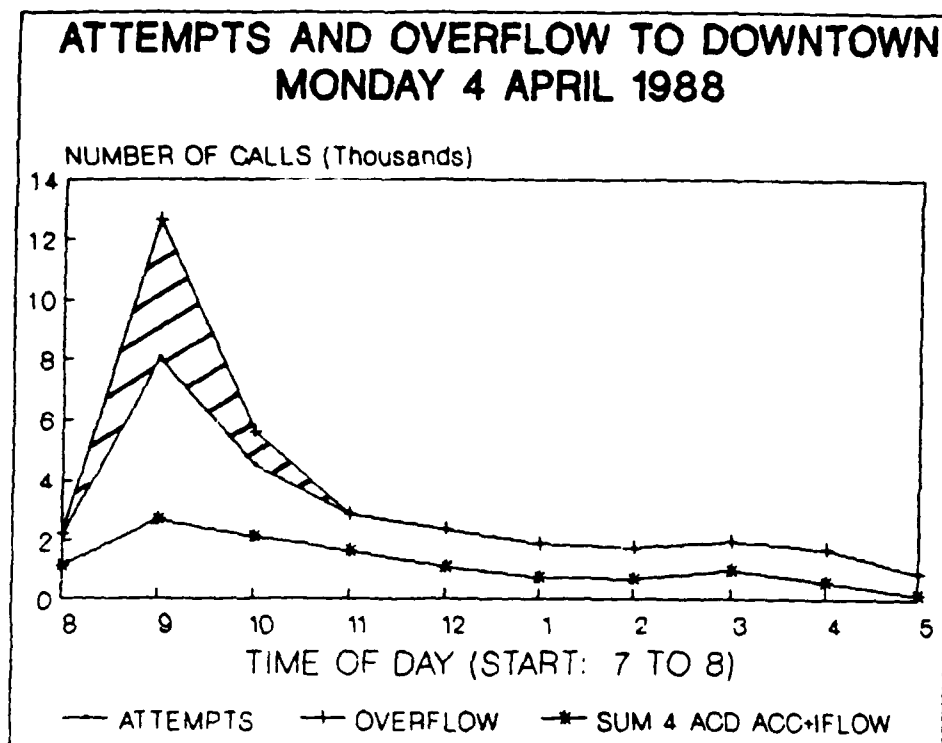


Figure 17. Downtown Minus ACD Traffic

From the ACD report on 4 April it was observed that once callers got into the system at the GYN site, they waited on average close to 14 minutes (833 seconds) to be answered. Those callers who reneged at the GYN site waited on average about 5 minutes (303 seconds) before reneging. The clerks on the GYN phones were often interrupted by a throng of walk-in patients demanding priority in booking appointments or otherwise wanting the attention of the phone clerks (52). The calls accepted graph for GYN in Figure 14 shows that from 7 to 8 A.M. over 200 attempts were made to reach the

GYN appointment phone most likely trying to game the clerks who sometimes open the phones a minute or so before 0800. Figure 15 shows the traffic for PEDS which has both a morning and afternoon peak.

The actual number of calls denied access is different for every graph based on scaling, but the percentage of denied calls is easy to estimate from the pictures. Figure 16 shows the combination of these four traffic streams, and GYN is clearly the dominant source of ACD traffic on the busiest day of the month. Figure 17 maps the combined traffic of the four ACD sites which book on Monday onto the downtown switch demand and overflow traffic for WHMC on 4 April. It is hypothesized that the call and time congestion due to the ACD queues booking appointments that morning was responsible for much of the overflow experienced by the downtown switch. Though the effect of redialing was impossible to estimate, this hypothesis was tested in the simulation experiments in Chapter VI.

Demand Traffic Conclusions. Peaks in interflow point out a major source of denied access and signal the need for additional servers, or a change in booking policy, to affect the characteristic demand pattern to allow greater access. Increasing the queue size at the ACD sites will not result in any better access, since no change in the system has yet to affect the number of calls answered. Longer times on-hold would increase the time congestion of the system which

would lead to more overflow. Interflow is the best indicator of demand even with the over-estimation due to redialing. Seasonality of demand within and between months, and on a yearly basis, should be measured using the interflow statistic. Since traffic patterns show most months are similar, since the first full workweek is unique in each month, since Monday of that week is the busiest day, and since 8 to 9 A.M. of that Monday is the busiest hour, investigation of this hour on 4 April 1988 seems reasonable as a focus for modeling telephone traffic at WHMC.

Observations made in the reduction of the data had a great impact on the research. The choice of the busiest hour of the month as 0800 to 0900 on Monday of the first full workweek of the month was confirmed. The usefulness of historical data to forecast calls accepted was shown, and the interflow measure was revealed as the best indicator of demand. The value of using the ACD data as feedback to look for changes due to policy changes was also proven in the case of the GYN ACD site. This evidence in support of assumptions made about the demand patterns to the ACD sites will give more strength to the evaluation of hypotheses tested in the simulation model.

V. Methodology

Overview

The processes of getting telephone access to WHMC and queuing behavior at the ten ACD sites embedded in the WHMC system were so complex, and the available data was so sparse and riddled with variation, that the best approximation to the WHMC phone system was chosen to be a simulation model. Analytical queuing and Markov process models were too simplistic to capture the activity in the WHMC telephone system, and the appropriate traffic and service conditions did not exist in order to use formulae from the telephone industry. The simulation technique allowed for random activities of various magnitudes to interact in a fashion similar to the in-bound and out-bound streams of calls at WHMC and its busiest appointment sites. The SLAM (53) simulation language was used to construct a model of telephone traffic for the general system and the ten ACD sites for the busiest hour of the month (Appendix B).

The method chosen was evaluated for merit and then used to fulfill research objectives. With an idea of the model's and the data's weaknesses, experiments were conducted on several control variables (the number of trunks, the size of the queue at the ACD sites, additional servers, and arrival

streams) to observe the impact on system overflow, and the number of calls answered across the ACD sites. The verification and validation stages of the research are below, while the findings of the experiments are presented in Chapter VI.

Modeling

The greatest effort was made to understand the true nature of the telephone system and caller behavior, while staying close to the context of the operational problem at WHMC. The common problem identified in the literature review was a trade-off between the level of service to the customer and the cost of manpower and equipment of the system. This is not the true nature at WHMC. The military medical community runs a not-for-profit agency regardless of how business-like some would try to make it look. The level of service desired by callers is getting an appointment and not just being answered. The clearinghouse activity and counselling service appointment clerks provide are important to caller satisfaction, but evaluating the impact on caller perception of this interaction is out of the scope of this effort. Furthermore, overhaul of the current appointment phone configuration is not desired for various reasons as alluded to in the extended background in Chapter II. Simulation modeling can describe the processes and

activities in the WHMC phone system to highlight chokepoints; no other technique fits the situation at WHMC well enough.

The telephone traffic and caller behavior into the phone system at WHMC and appointment centers are best represented as a two stage, heterogeneous, multi-site, multi-server, loss/delay queuing system with redialing. No queuing method allows for renegeing and redialing behavior, nor can such a method easily entertain the notion of many dependent arrival streams competing for common resources; therefore, simulation was used to emulate the system. Relevant parameters were: 1) the interarrival rate of calls; 2) the number of lines available to the entire hospital; 3) the number of waiting places allowed in the queue at the telephone appointment center; 4) the rate at which callers will renege from the queue should they not receive service within some time limit; 5) the service rate for processing calls; 6) the number of available servers; and, 7) the time servers lose due to interruptions in, or vacations from, their telephone answering tasks. Many streams of in-coming calls compete for the lines available to the hospital, and there can be significant blocking to entering the system during the busiest times of the work week.

Assumptions. The following is a list of assumptions used in this study:

1) Data from the ACD system contains accurate peg-counts of call activity, and also contains accurate call processing data for service times (see Chapter 4).

2) Future teletraffic demand can be estimated from historical data.

3) Non-ACD traffic is answered if it reaches the WHMC switch.

4) Regardless of the fact that the distributions for arrivals, reneging, and service times are undiscernable, the exponential distribution can be used as a conservative surrogate since the first moment of each of these variables is estimable from the data.

5) Actions which improve service during the heaviest system traffic during the busy hour will also help to improve service during the other 39 hours of the workweek.

6) ACD sites which do not book on Monday can take similar actions as those which do, in order to improve service.

The Structure of the Simulation. The simulation model links essentially identical processes into a competitive framework. The process is the arrival, waiting, and leaving of calls. The competition is two-fold, arriving telephone traffic streams (the non-ACD and the ten ACD sites) compete to seize one of the ninety commercial trunks allowing access to the WHMC switch. Then, the ACD traffic competes at the individual sites to join a queue for the appointment

clerk(s) (the implicit third queue of available appointments is not a part of this study). As shown much earlier in the system diagram (Figure 2), failure to reach the WHMC switch is counted as overflow traffic, while failure to join an appointment queue after siezing a trunk is counted as interflow traffic. Reneging behavior is permitted. The renege event is scheduled at the time the call queues, and is executed if service has not begun by the time the call would renege if it waited long enough. Redialing is ignored and left as inflated demand as in the actual data and as in the simulation models in the literature review. Much time was spent reviewing techniques in the literature for defining the redialing behavior of the ACD sites and the non-ACD traffic, but no solution was found. Queue statistics are collected for each ACD site, and utilization of the trunks is also gathered. The period of concern is one hour (the busiest hour), simulated as sixty minutes. The structure of the simulation is simple, and code for the model and the Fortran subroutines are in Appendix B.

Parameter Calculation. Model input consisted of separate estimates for each ACD site's arrivals, reneges, service times, post-call processing times, and number of clerks, and non-ACD arrivals and service times. The formulas for calculating these variables directly from the raw data from the ACD reports are summarized in Table I, while a narration on creating these parameters follows.

Interarrival times for each ACD site were the inverse of the total of calls accepted and interflow for the hour divided by sixty (minutes per hour) to represent interarrival times in terms of minutes. Renege, service, and post-call processing parameters were averages of data directly from the ACD reports for similar hours of similar days (as previously described) and divided by sixty (seconds per minute) to convert the data collected in seconds to minutes. The number of servers was the average of servers used (position opened and serving calls) at each site for the similar time periods of the data rounded to the nearest integer. Non-ACD data parameters were backed-out from the data of the April 1988 downtown switch traffic study using that week's ACD reports. The ACD data was subtracted from the downtown switch data in order to represent the rest of the WHMC telephone traffic (non-ACD).

Table I. Input Parameter Calculation Equations

Arrivals = (Calls Accepted + Interflow)/60	(per minute)
Interarrival Time = 1/Arrivals	(in minutes)
Renege Rate = Average Time to Abandon/60	(in minutes)
DCP (service rate) = Average DCP/60	(in minutes)
PCP (interruption) = Average PCP/60	(in minutes)

Below, Table II lists the input parameters for the baseline model. The exponential distribution was used for generating all the random numbers as a conservative projection, since when only the average for a distribution

is ascertainable, the exponential distribution provides a conservative distribution for the variable concerned. The parameters come from averages of the five common hours of data for the busiest hour of the five months of comparable data available. The parameters have not been manipulated to make the model perform like the actual data. Instead, the model embraces the problems of the real data for ease of use and replication should WHMC, or any other organization, desire to use this model.

Table II. Baseline Input Parameters for the Model

ACD Site	Number of Servers	Inter-Arrival Rate	Reneged Rate	DCP	PCP
PCARE	2	0.4140	0.8633	1.3600	0.3667
ORTH	1	0.1500	1.7833	2.4333	0.2933
GYN	1	0.0360	3.4167	2.0000	0.4233
IMED	1	1.9260	0.5667	1.7967	0.2200
OPHTH	1	1.7460	0.6100	1.1833	0.0333
OB	1	2.4180	2.2000	1.2117	2.8833
PEDS	2	0.8700	2.5833	1.8667	1.7667
DERM	1	1.3620	2.1600	2.0667	0.4667
GSURG	1	1.7520	1.1367	0.8667	0.5000
CARD	1	2.7000	1.3833	2.2167	1.7033
NONACD	BIG	0.01156	15.00	1.8333	N/A

Relating parameters and ACD data through the equations in Table I is a simple task which will be illustrated using the parameters for the PCARE site in Table II. The total of the average number of calls accepted and interflow for PCARE during the busy hour was 144.9, 70.2 calls accepted and 74.7 calls interflow ($1/(144.9/60)=0.4140$), so the interarrival

parameter is 0.4140. The renege rate was calculated from the average time to abandon statistic of 51.8 seconds ($51.8/60=0.8633$). Similarly, the DCP and PCP rates were found from 81.6 and 22 seconds respectively.

Verification and Sensitivity

The simplest verification permitted was to eliminate overflow, renegeing, and post-call processing, and then compare site results with analytic queuing results. A spreadsheet was programmed with the equations for the M/M/s/k queue (31) allowing for up to four servers and for a queue capacity between three to ten. Comparison of analytic to simulated queues for each ACD site indicated that when no interference other than the queue capacity was involved, the simulation reproduced accurate queuing behavior. Therefore, the structure of the model was verified to be correct. Hence, only poor data and the effects of allowing overflow, renegeing, and post call processing to occur in the queues could account for peculiarities in the output of the simulation.

The ACD Site Module. Detailed sensitivity analysis using Response Surface Methodology (RSM) (6) was conducted on the module for an individual ACD site for simplicity with the assumption that the nine identical modules would behave similarly in the larger model. The experiment consisted of

varying the levels of five variables in the simulation of one ACD site and running the model a number of times to achieve an average time-on-hold. The five variables were: 1) the number of lines available for the hospital (TRUNKS); 2) the number of waiting spaces at the queue (STACK); 3) the number of servers (SERVERS); 4) the rate of reneging from the queue (RENEGE); and, 5) the interarrival rate of calls (INTARR). The number of TRUNKS was set to approximate the conditions observed in the more complex model, so that some blocking of calls could occur at a high intensity of traffic. The number of STACK was set to six which was the policy observed in the hospital. Each remaining variable had a nominal level fixed using actual system data from the busiest hour of the busiest day of the month of April 1988 at the FCARE site. The busiest time was used in order to give management the "worst case" look of demand for service. The actual parameters are not presented in order to not give the false impression of having a site specific metamodel. The RSM technique was used solely to examine the repeated module of the simulation model for sensitivity to input parameters, and number of servers and the renege rate were the only significant variables for the linear model developed through the RSM technique.

The procedure followed in the experiment was as follows. High and low values (+/- 10% for rate variables, 3 and 1 for the number of servers, and 10 and 4 for the queue

length) were chosen about the nominal values (from the actual data for PCARE as above) for each of the variables. Parameters were then normalized to values of plus and minus one. A 2^5 full factorial experimental design on the five factors was chosen to give the best coverage of the experimental region. The simulation model was run ten times at each of the 32 possible variable settings (design points). The response for time-on-hold was averaged across the ten runs to get a response for that design point. Variance reduction techniques for the simulation were employed by using the assignment rule for common random number streams (56). Additionally, 50 runs were made at the nominal values, or center point, of the experimental region and after averaging as before, five observations were obtained at the center point.

A linear metamodel was fit through the full factorial experimental design of a simulation model of a loss/delay queue with multiple servers, reneging of callers, with a queue capacity with nominal length of six at the busiest hour of the month. There were four significant terms for interpretation: 1) a positive intercept, which speaks for itself; 2) a negative interaction with the number of servers, so that increasing the number of servers decreased time-on-hold; 3) a positive effect from the renege rate, so that when callers reneged after holding longer the average time-on-hold increased; and, 4) a negative two-way effect

between SERVERS and RENEGE. The percent of the variation in the data explained by the model, the R-square statistic, was 0.9797 and the F statistic was over 450, but the model underestimated the actual time on-hold for the ACD site (the actual data was four times the predicted).

The linear model failed a lack of fit test as expected, but still gave valuable insight into the problem (there was negligible error of 5 seconds between the linear model and the center point of the experimental hypercube). When measuring time-on-hold for a queue it is common knowledge that the curve is close to flat at low traffic intensities and rises exponentially to a very long time-on-hold at high intensities (approaching one). A second degree RSM model is typically used. The busy hour of the busy day scenario had an arrival rate much greater than the service rate so that the intensity was greater than one. The most significant findings of the RSM study were that the module was found to be sensitive to the renege distribution, and that adding servers was the only way to decrease the time on-hold statistic. The more complex model was needed in order to study the effect of the number of trunk lines on the traffic to other parts of the hospital, and to see the congestion and competition effect of the arrival streams at various ACD sites.

Warm-up Periods for the Model. The actual traffic for the busy hour is highest at the start of the hour, and ACD

sites already have callers in their queues. The model used hourly averages from the ACD system, so it could not capture the different traffic intensities between the first and last half of the hour, but it was possible to warm-up the simulation by letting it run a while before beginning to collect statistics. Different warm-up periods were compared to see the effect on the simulation results. The simulation was allowed to run for five, ten, and fifteen minutes before clearing the registers to collect statistics for the hour. There was no significant difference in output using the larger start-up times, so five minutes was the standard. The ACD site demand and service data graphed earlier in the paper clearly indicates that there is no steady state for this period of operation. The high intensities of traffic and blocking at the downtown switch and at the ACD queues combined with reneging make the system chaotic.

Validation

Two validation efforts were conducted based on system benchmarks and face validity from knowledge of the system. Benchmarks for the model came from the data from the downtown switch study, averaged ACD data peg-counts, and from an inferred upperlimit on the number of calls served per clerk. The problems of statistical validation for multiple response simulations is summarized by Shannon

(58:294-295), and no concrete single measure for combining and weighting all the different outputs of the non-ACD and ACD queues was contrived. Instead, server capacity was linked to the number of calls answered, and throughput, the number of calls answered in the hour, was used as a simple signal of the models' success in approximating the WHMC phone system. Interflow and overflow were also benchmarks used to validate the model.

Time on-hold. The traditional queuing measure of time on-hold was useless as a validation check due to the inaccuracy and variability of the original data and the model's sensitivity to the reneging distribution. In every case, the model reports significantly lower times on-hold than actual ACD data for average speed of answer (ASA). The lowest average time on-hold for any ACD site in the actual data for the busiest hour was a minute and a half (CARD) with the median time on-hold for all ACD sites being over three minutes. The baseline simulation model reported times-on-hold as low as 10 seconds. However, the model did show the longest times on-hold belonged to the same ACD sites as in the actual data (GYN and ORTHP), but even these results were barely fifty percent of the time on-hold values of the actual data.

The time on-hold at a single ACD site was found to be very sensitive to the renege rate in the RSM study, and the inability to model clerk nonperformance from the ACD data

added variability. This made it impossible to achieve holding times similar to the actual ACD data. Using the average-speed-of-answer (ASA) statistic from the ACD reports as a measure of how long callers are willing to hold might work out better than using the average time-to-abandon statistic for getting closer estimates of time on-hold. The model remains sensitive to reneging (just as the models reviewed in the literature). While the time on-hold criterion of the individual sites would be a boon to the face validity of the model, it would be difficult to use a single, combined time on-hold measure across all ten ACD sites for evaluating the system model.

Reneging and Interflow. When compared to the actual busy hour data, the model underestimated the number of callers who reneged and the number of callers who got the "all circuits busy" message, interflow, for all of the ACD sites. However, when the model showed reneging and interflow, it did so at the same ACD sites where reneging and interflow were the most severe in the actual data. The renege and interflow measures also followed the trends of common knowledge of the system when the queue size and the number of trunks were varied. When the queue was set higher, the number of reneges increased while interflow decreased (allowing more calls to queue results in fewer busy signals, but induces more calls to renege due to long hold times). Table III shows the effect of varying the time

to renege on the number of calls answered and on system overflow for the baseline model. The short time for renegeing was half a minute, while the long time to renege was 20 minutes. When more calls wait in the queue for longer periods, fewer lines remain to accept new calls and overflow rises due to the time congestion of the lines. When calls release the lines quickly, more calls can use the lines and overflow decreases. However, this had little effect on the number of calls answered, since in heavy traffic with the queue set at 6 there would typically be a call awaiting the clerk regardless of the waiting time in the queue.

Table III. Effect of Time to Renege on Calls Answered and System Overflow

	ACTUAL DATA	BASELINE RESULTS	SHORT TIME	LONG TIME
PCARE	67	51	50	52
ORTHP	19	19	20	21
GYN	30	24	21	22
IMED	29	10	11	13
OPTHL	17	14	12	14
OB	5	8	9	8
PEDS	30	22	21	22
DERM	24	13	14	12
GSURG	26	12	12	13
CARD	12	6	7	8
TOTAL ACD	259	179	177	185
OVERFLOW	4639	4307	4028	4689

Calls Answered and Overflow. The benchmarks for evaluation of the model were the average total number of calls answered at the ACD sites, and the average overflow of the system. These output variables were much more useful for comparing the simulation to the actual ACD data than the time on-hold measure. From a face validity standpoint, it was positive feedback to see the peg-counts in the actual data for these variables compared favorably with the baseline simulation output (Table III). From the standpoint of the objective of the study, trying to help maximize the throughput of the system while minimizing blocking at the downtown switch, these measures seemed most amenable.

For face validity, the model's output was low for the number of calls answered statistic of the ACD sites, but the overflow statistic for the system was closely approximated (Table III for comparison of the baseline model with the actual data). Closely fitting the model to all the parameters of the actual data was not the priority for evaluating or validating the model with all of the difficulties surrounding the simulation: 1) the variability in the (just) five sets of available data used for the parameters; 2) the sensitivity of the ACD site modules to the renege distribution; and, 3) the hourly averaged data versus the actual half hourly changes in the nature of the traffic. Instead, a close fit or reasonable numbers (considering knowledge about service rate upperbounds) which

were consistent over the different ACD sites were the criterion for accepting face validity of the model. For every ACD site, the baseline model produced a number of calls served that was either near the actual number of calls served or was below the upperlimit for servers. The model also produced an overflow statistic much like the actual data (4307 to 4639) as already shown above. Both of these measures are used in the next chapter for comparison and evaluation of experiments conducted on the model to increase throughput and decrease overflow.

Performance Measurement

The chief concern of management is the impact that perceived accessibility to care has on the military community's perception of health care quality. Any change which helps more calls to be answered in the busiest hour, should also carry over to the more relaxed system in the other hours of the week. Though no quantitative link has been established between caller perception and any measurable ACD statistic, many indices might be useful as surrogate gauges for access such as: 1) the number of calls answered; 2) the percentage of calls answered versus blocked or reneged; 3) the number of callers who get an appointment; 4) the time it took to reach the appointment clerk; or, many other types of indices. For the current

effort, access is be defined in terms of the number of callers who get to speak with an appointment clerk (calls answered) for the ACD system and by the amount of overflow at the downtown switch for the WHMC telephone system. The goal is to maximize the system throughput, the number of calls answered at each ACD site, while minimizing overflow to the downtown switch.

Experiments Conducted

To achieve the goal of this study it was necessary to look at alternatives through simulation. These alternatives considered changes to the phone system hardware, manpower, service policies, and scheduling of the ACD sites. The two hardware experiments were, changing the size of the queue at the ACD sites, and changing the number of commercial trunks leased. Under manpower, the question of where to put an additional server was addressed. Service standard experiments covered reducing the variability of service times by implementing a policy of no interruptions of the clerk and increasing the speed of service to an upper limit (different limits were used for computerized versus logbook systems). Finally, an experiment in scheduling addressed the possibility of moving all of the Monday ACD traffic to the similar time block on Wednesday morning to see the effect of not opening appointment books on the busiest day of the week. Thus, clerk time would be freed at these ACD

sites for Monday morning duties other than booking appointments by phone.

This model could be useful for testing the effects of consolidating ACD sites. Arrival streams could be combined and servers could be pooled in order to evaluate improved system throughput and savings in manpower. Many of the simulated clerks (ACD sites not booking on Monday) were operating at low to moderate utilization levels (30 to 60%). The actual data for call arrivals during the busy hour shows that six of the ten sites had 35 or fewer arrivals in the hour. This scenario was not explored, since radical change was not permitted, and since insufficient information was available to determine which sites were similar enough in required medical knowledge to allow the training and pooling of clerks. However, as noted in the literature review, consolidating service centers for the single task of making appointments would make that task more efficient and contribute to system throughput.

Two experiments that would have been of interest were untestable. The first was the idea of using answering machines after all appointments were booked (as the OPTH L ACD site did over the course of this study) to quickly answer callers with the appointment information and to free clerks for other tasks. The second was implementing a policy of a rolling appointment window which would open every day (also as in OPTH L) for that same day of the week

two to six weeks in the future depending on the site. Going to this booking policy would change the characteristic demand at the ACD sites, and the demand distributions this study uses from the historical data would not apply.

VI. EXPERIMENTATION AND ANALYSIS

Overview

Four areas were identified for improving the access to WHMC as defined by the number of calls answered and the amount of overflow. Hardware, manpower, service policies, and scheduling are all areas which can be manipulated to some degree at WHMC in order to affect the access to the ACD appointment sites and to the hospital in general. Experiments were conducted with the simulation model to evaluate how the calls answered and overflow criteria varied with changes to the controllable factors. Results indicated that without changing the schedule the best way to improve access would be to shorten the queues at each ACD site, add manpower to the busiest sites, and increase server productivity by not allowing interruptions and by computerizing the appointment making task. If a change to the schedule of booking days is implemented to unload Monday traffic, the new Monday traffic will experience less overflow and the new appointment day(s) will see no overflow. Remembering the simulation model could not account for redialing, one might expect the true effect on the overflow for the new Monday traffic could be much greater than simulated due to fewer redialing ACD callers in the actual system.

Running the Experiments

The simulation was repeated at specified variable levels for ten runs and the data for calls answered and overflow were averaged across these ten repetitions. The number of repetitions chosen came from observing the operation of the model (58:297). The model ran quickly on the computer (about a two minute actual delay). Up to fifty repetitions were originally used for an experiment as what seemed a large enough sample size since a priori statistical parameters were not applicable for determining run size. Through observation it was seen that the averaged data over the fifty runs rounded to the nearest integer was the same as the averaged data over ten runs, so ten runs became the run size. A number below ten runs might prove to be as sufficient, but this was not investigated. The observed direction and magnitudes of the trends in the output variables of calls served and overflow in relation to the baseline model were the basis for acknowledging an improvement or fall in accessibility.

Changes to Hardware

Two changes in the telephone hardware are possible. The number of commercial trunks leased and the length of the

queues at the ACD sites are both controlled by the communications branch of WHMC. Trunks cost in the area of fifty dollars each a month to lease, while setting the mutual queue size at the ACD sites is easily done by pushing a few buttons. The number of commercial lines was set at many different levels: 1) the pre-November 1987 level of 60 lines with a veritable infinite queue at each ACD site; 2) the baseline model of 90 lines; and, 3) at various other levels to explore the experimental region. Table IV contains the results of the experiment. The system was in the worst condition during the busy hours in early 1987, but even after shortening the queue to 6 (as in the baseline model to cut down the phone bills of callers), the system with only 60 lines still has significant blocking. The increase from 60 to 90 lines shows that the value of the 30 lines added by WHMC communications lies in access to the non-ACD calls to WHMC, and has only a slight effect on ACD calls served. As the number of lines is increased, overflow decreases and the number of calls answered increases. The number of calls answered at any particular site tends to reach an upper limit which cannot exceed that clerk's service rate. It is possible to buy enough lines so that no overflow occurs, and when the model was given 200 lines to see how many lines it would use, the most trunks in use at any time was about 140. Although this might be a desirable way to deal with overflow, other, less costly, alternatives

can be employed to lessen overflow at the busiest hour of the month without wasting so many lines when the traffic is lower on the other 38 hours of the workweek.

Table IV. Effect of the Number of Commercial Lines on Calls Answered and System Overflow

	< NOV '87						
	ACTUAL	QUEUE=BIG		BASELINE			
	LINE=90	LINE=60	LINE=60	LINE=90	LINE=100	LINE=120	LINE=200
PCARE	67	34	50	51	52	64	65
ORTHOP	19	18	19	19	22	24	25
GYN	30	29	26	24	25	25	25
IMED	29	7	8	10	12	18	20
OPHTH	17	5	9	14	15	19	19
OB	5	5	5	8	8	11	12
PEDS	30	14	20	22	25	34	38
DERM	24	7	11	13	14	14	16
GSURG	26	5	9	12	15	19	20
CARD	12	5	5	6	8	10	12
TOTAL	259	129	162	179	196	238	252
OVERFLOW	4639	6409	5743	4307	3499	2102	0

The length of the queue at the ACD sites was varied between low and high extremes to observe its effect on access to care. The results are tabulated in Table V in order of increasing queue length. The length of the queue had little effect on the number of calls answered until the time congestion caused by more calls holding telephone resources longer made system overflow and reneging increase. The interflow (not shown) did increase as expected for the smaller holding area and decrease for the larger queue size. By decreasing the number of people who could be on-hold at the ACD sites, the overflow decreased since fewer callers

could seize lines for shorter times. Access to care, based on overflow, decreases at higher queue lengths while no benefit in the number of calls answered is realized.

Table V. Effect of Queue Size on Calls Answered and System Overflow

	ACTUAL QUEUE=6	QUEUE=2	QUEUE=4	BASELINE QUEUE=6	QUEUE=10	QUEUE=60
PCARE	67	64	49	51	46	42
ORTHOP	19	18	21	19	21	18
GYN	30	25	25	24	25	25
IMED	29	13	11	10	11	9
OPHTH	17	13	15	14	13	9
OB	5	6	9	8	8	9
PEDS	30	22	24	22	22	22
DERM	24	12	14	13	13	10
GSURG	26	14	16	12	12	15
CARD	12	6	7	6	6	7
TOTAL	259	193	191	179	177	166
OVERFLOW	4639	3969	4210	4307	4558	5359

Choosing a queue length to accommodate callers to the ACD appointment sites is a special issue. As stated in earlier chapters, all ACD sites are run by common software which only allows one queue size to be set for all the sites. If the queue size were set to a very large number as before November 1987, callers to the busy ACD sites could take over the system (i.e., 10 calls at the 10 sites is more than the 90 trunks). The callers arriving at the end of the line would likely renege due to the long hold time, but if they held-on, their wait would be large (and so would their long distance phone bill). Even worse, they might wait a

very long time and still not get an appointment. The policy for keeping the queue long allows the callers with the ability to hold the longest to get through to the clerk, but does not increase the number served. If the queue length is set low, many callers get the "all circuits busy" message (interflow increases), but no one waits for very long on average. Still, the same number of calls get answered, but more quickly than before.

Any queue length policy in the excess demand situation of the busy hour simply governs the length of wait and the amount of interflow (redialing attempts included), since having longer queue lengths does not affect the number of calls answered. If the $M/M/1/K$ queue is used as the approximation for a site with a service rate of 25 calls an hour, then the wait in queue with length 4 would be around 6 minutes in a very high demand environment (intensity of 3, e.g. GYN and DRTHP for busy hour), and around 11 minutes for a queue length of 6. For the two server approximation under similar service and demand using the $M/M/2/K$ queue, the waits for K equal to 4, 6, and 10 are about 2, 4, and 9 minutes respectively. Using calls accepted is a false surrogate for access; the calls into the queue must be answered to truly achieve access, so quickly serving the callers who get in so they do not renege is better than worrying about the already high interflow for the sites during their busiest periods when considering access.

Reneging from the queues with one server was twice as high when the queue length was set at 10 versus at 4. Since eight of the ten ACD sites historically operate with only one server at the busy hour, a queue length should be chosen for the one server case, and a conscious decision to put an upper limit on waiting time can decide the appropriate queue length. Reducing the queue length to eliminate renegeing can be done adaptively, by lowering the queue length and observing the renegeing phenomenon. Once the renegeing is controlled, any further reduction of the queue length will decrease the time on-hold and can again be controlled adaptively to set the average time on-hold for the callers, so long as the clerks are operating efficiently. On the other hand, the longer the queue, the fewer busy signals, so controlling the queue length so that renegeing is just eliminated is the best condition at which to operate versus doing away with the queuing structure and going to a one or two line office phone.

Data from the ACD system can be checked to observe the progress of actual experimentation with the queue length on wait time and interflow. The ACD reports are printed on the hour and the queue length is easy to change by a push of a few buttons. The averaged actual data for the simulation model showed intensities over three for the two sites mentioned above, but these two sites also contributed almost all of the interflow for the hour. Each site has its own

demand intensity during its busy hour on the day it books appointments, but all sites have more calls accepted than answered for the busy hour of the busy day of the system in this study. The other six ACD sites with only one server still experience intensities of traffic greater than one (about 1.4), but with negligible interflow. The callers to these six sites can expect to wait about 8 minutes using the M/M/1/K approximation with the queue size at 6. The average wait from the actual data for the busiest hour of Monday morning varies by site, but for sites answering at least 25 calls that hour the average speed of answer was a little over 4 minutes. This sheds doubt on the usefulness of the M/M/c/K queue for approximation of these ACD sites. A shorter queue would lower the wait time so that renegeing would lessen. Interflow was negligible at seven sites and will not likely increase greatly for the system if the queue is shortened.

Under the definition of access used in this study, the best access is achieved when calls accepted equals calls answered. Information on the renegeing distributions for the different ACD sites is too poor to use in establishing the best queue size at this time. Trials should be conducted with the actual ACD system to lower the queue size from 6 until renegeing behavior is negligible. In essence, those who renege are just interflow which waste more time congesting the lines that otherwise could be serving a call

that would get through to service. This policy will assure that the calls which are answered get the fastest service; the others never make it to the clerk anyway.

Manpower

One sure way to increase the amount of calls which get answered is to have more people at work, assuming they do the task with some degree of efficiency. Personnel at WHMC feel 25 calls an hour is the amount of calls answered for an average clerk. Observed cases in the data have gone as high as 46 calls answered in an hour (8 to 9 A.M., Monday 1 August 1988, PCARE), but this was an outlier. A good yardstick might be 25 calls an hour on a manual system and up to 30 to 35 calls an hour when using a computerized appointment system. Hence, the computer becomes a force multiplier for appointment clerks, but people still have to be available to answer the phone. Getting more out of the people currently at the task is the subject of the next section, this one deals with how to allocate the people currently at the task and where to put any new help that may come along.

Staffing to meet demand is a great concept unless the situation is overconstrained. Additional manpower at WHMC is a precious commodity. If one new person were to be allocated upon the basis of aiding access to the medical

center through the telephone appointment center, the logical position to put them would be at the busiest site. However, this can be confounded by the day of the week, since each site is busiest on the day it books its own appointments. GYN is clearly the choice on the first Monday of the first full work-week of the month, but other ACD sites also need help that day, and every site needs help in dealing with peak demand periods. In the experiment titled GYN3 in Table VI below, two new clerks were added to the baseline simulation at the GYN site and one new server was added to the PCARE site to bring them both to three clerks. The calls answered at GYN went from 24 to 75, while the calls answered at PCARE only rose by 7 to 58 calls answered. The reason for the disparity is that from 0800 to 0900, GYN was under intense traffic and an average interflow around 700, while PCARE was answering nearly 90 percent of its calls with little interflow. Although the people calling PCARE found excellent service and the clerks had some leisure time at the busiest part of the week, many of the calls to GYN were still being blocked. The point is that by assigning the new servers to the busiest site(s), the throughput of the system is maximized. The new clerks at GYN averaged about 25 calls, while the new clerk at PCARE only accounted for 7 more calls answered. To explore what would happen should GYN get four helpers, either fulltime or augmentees, the GYN5 experiment in Table VI showed that the demand was

so great at the busiest hour at GYN that all five clerks would be near their upper limits for servicing calls (117/5 is about 23 calls each).

Table VI. Effect of Adding Servers to the Busiest Site

	ACTUAL	BASELINE	GYN3	GYN5
PCARE	67	51	58	52
ORTHP	19	19	22	21
GYN	30	25	75	117
IMED	29	10	11	9
OPTHL	17	14	12	15
OB	5	8	8	10
PEDS	30	22	24	25
DERM	24	13	14	13
GSURG	26	12	12	14
CARD	12	6	7	9
TOTAL ACD	259	179	243	285
OVERFLOW	4639	4307	4429	4478

One way to alleviate stress at all the sites and yet let each site maintain authority over its clerks would be for the managers of the sites to agree to loan clerks to one another to aid in serving peak periods of telephone traffic. This could work if the ACD sites had: 1) peak periods which did not coincide; 2) enough similar hardware (or log books) and space for augmented manpower to use; and, 3) a similar type of required medical and appointment knowledge so that an augmentee could quickly master the new site procedures for making appointments, giving information, and identifying

cases needing referral. The training currently evolving through the Ambulatory Care Initiative (ACI) insures all clerks have the proper courtesy training, and a member of the ACI cadre might be a good reference to help coordinate a voluntary matrix approach to pooling manpower during the demand peaks of different ACD sites. Although illustrating this over a week is beyond the capabilities of the parameter database of this research, sharing limited resources during high demand periods is one way to increase the amount of manpower within the current system so that access to each ACD site can be increased.

Service Policies

Service policies dictate the speed of service to the customer. Environment and process both can play a role in the speed of service. Environmental aspects in service speed include the physical space where the clerk operates and the type of equipment the clerk uses, while the process aspects of service are buried in the management actions of the ACD site. Environment can hinder service speed if the clerk is frequently interrupted by staff or walk-in patients while making appointments, or if the log book system used makes it very difficult and time consuming to reference appointment times using Social Security numbers as is most often done. The process of making appointments hinders

service speed when the site management is lax and the clerks are not as productive as they otherwise might be. The process is easily overlooked at WHMC for two reasons: 1) by the criteria of appointments booked as the most visible indicator of an appointment clerk's productivity all the available appointments are indeed being filled; and 2) no feedback mechanism to show the traffic to the individual ACD sites and clerk productivity is in place to be used, so site managers often are not aware of the traffic, but just that the demand is greater than the supply of appointments. The access criteria of calls answered in this study would require a feedback loop of a performance indicator on calls answered to help clerks gauge their productivity and to get site management involved in the process of answering calls at each site.

To model the benefit of changing service factors the simulation model was run to reflect having no interruption of the clerks during the hour, and to also allow for having all sites on a computer system. A service rate of 35 calls answered an hour was then set for each clerk as an absolute upper limit on the number of calls they could possibly answer. To simulate no interruptions, the post call processing time was set to zero to account for few to no interruptions. This would not account for interruptions while the clerk was on the phone with a caller. When total ACD calls answered for the model with no interruptions is

compared to the baseline, the amount increased on average by 34 calls to 213 calls answered in the hour (Table VII). ACD sites previously experiencing interruptions in their phone tasks were able to complete more calls when left to their tasks.

Table VII. Effect of Speed of Service on
Calls Answered and System Overflow

	ACTUAL	BASE- LINE	NO INTER- RUPT	ALL USE COMPUTER
PCARE	67	51	53	52
ORTHOP	19	19	21	34
GYN	30	24	30	33
IMED	29	10	11	11
OPHTH	17	14	14	15
OB	5	8	11	11
FEDS	30	22	30	30
DERM	24	12	16	16
GSURG	26	12	15	11
CARD	12	6	8	8
TOTAL ACD	259	179	213	221
OVERFLOW	4639	4307	4372	4386

The service rate was set uniformly at 35 calls answered an hour for evaluating the effect of computerizing all the ACD sites on calls answered. The upper limit of 35 was used with post call processing set to zero to reflect an environment of isolated appointment clerks without interruptions who had the best equipment available to do their tasks. Table VII indicates that the total ACD calls answered increases to 221, 42 over the baseline model and slightly more than just the policy of no interruptions. The

computer systems however add other aspects of productivity not seen in calls answered. Not only do computerized appointment systems allow quickly finding patient and appointment data, thus lowering the variance of individual service times, but they track and calculate routine statistics on all the physicians and patients using the facilities which would otherwise be done manually in a very cumbersome and time consuming fashion. The computer system frees the clerks to do more nontrivial tasks once telephone traffic dies down.

Scheduling

Since the system experiences greater demand on Monday than any other day of the week, the idea of scheduling appointment availability at ACD sites so as not to be busy making appointments on Monday mornings came naturally in an effort to balance the demand on telephone resources, and on clerks across the week. The simulation was not designed to test scheduling feasibilities; however, it was possible to experiment with a scenario of decreasing the amount of traffic to WHMC on Monday morning. The ACD sites contributing the most to Monday morning demand were the ones which booked appointments that morning (PCARE, GYN, ORTHP, and PEDS). The experiment consisted of two elements: 1) the effect on Monday morning overflow of decreasing the

traffic; and, 2) the behavior of the system on the new time these four appointment sites would open for business. Wednesday was used as the new day for booking appointments since it was a slow day for all the sites and since no other ACD site used this day. Wednesday parameters for arrivals, renegeing, and service were added to the Monday model to substitute for the original parameters of booking on Monday in order to represent the reduced traffic. In addition, the case of only decreasing traffic to the GYN site was explored, since in the real system a change in GYN traffic in August 1988 reduced demand intensity and interflow at the site for the busiest hour of the month.

Table VIII. Effect of Schedule Changes on Calls Answered and System Overflow

	ACTUAL	BASE LINE	MON - GYN	MON - 4ACD	MON - 4ACD+	BOOK WED
PCARE	67	51	47	14	19	60
ORTHOP	19	19	20	11	16	20
GYN	30	24	24	20	19	26
IMED	29	10	12	11	15	17
OPHTH	17	14	13	15	16	21
OB	5	8	7	8	8	11
PEDS	30	25	23	15	16	33
DERM	24	12	15	13	16	20
GSURG	26	12	14	16	17	25
CARD	12	5	7	8	9	11
TOTAL ACD	259	180	182	131	151	244
OVERFLOW	4369	4307	3853	3261	1761	0

Table VIII contains the data for the experiments on the scheduling of appointments with the baseline simulation of

the busy hour. Without the high demand traffic parameters for the four ACD sites which book on Monday (MON - 4ACD), the total ACD calls answered is much lower since demand is much lower, and the overflow dropped by over a thousand from the baseline of 3261. When nothing but changing the GYN traffic (cut in half) is done, overflow still drops about 450 to 3853 with no real difference in the number of calls answered by the ACD sites (Table VIII, column MON - GYN). The test case for moving all ACD booking off Monday was supplemented (Table VIII, MON - 4ACD+) in another experiment by hypothesizing the ACD call and time congestion might block up to twenty percent of the non-ACD traffic, thus causing redialing and inflating non-ACD demand. When non-ACD traffic was reduced by twenty percent with no ACD booking activity on Monday, the overflow dropped to only 1761, about 2.5 times less than the baseline. Eliminating peaks of excess demand by changes in scheduling can improve access by decreasing overflow while not affecting calls answered.

The effect on the ACD site access by moving all the booking days off of Monday was also significant. By using the Wednesday non-ACD traffic parameter in the Monday morning model, the simulation put typical Monday morning ACD traffic competing against a much lower demand from non-ACD traffic for the commercial trunks. The effect on access was that there was zero overflow, and calls answered rose to 244

from the baseline of 180 (Table VIII, BOOK WED). Thus, the overflow of the system on Monday mornings can affect the number of ACD calls answered by blocking calls destined for an ACD site which otherwise might have had a good chance of being answered before reneging.

The most remarkable finding from using the Wednesday non-ACD traffic parameter was that it was the first time in all the experimental runs when trunks were not utilized nearly 100 percent. In fact, only 50 trunks were required on average to handle the Wednesday traffic even with the four ACD sites now using Wednesday to book appointments. The case for moving traffic off of Monday is readily apparent from an access to care standpoint, as well as from a system sizing viewpoint. Better access and less expense for leasing trunks can come from balancing traffic by smoothing Monday morning demand.

Further Research Areas

Several issues for describing access to ACD sites across the week and for influencing caller perception of accessibility remain unexamined. While it is an implicit assumption that trends and observations of the busy hour will be useful to all ACD sites, it must wait to be seen how the six sites which open for booking on days other than Monday will actually react. It can be expected that peaks

in demand and calls answered at these sites will react the same, but since overflow only occurs on Monday mornings, no effect of overflow on calls answered will be found. The extent of redialing due to interflow and/or overflow is still an unknown, but redialing does inflate demand estimates for the system. This larger demand causes overprovisioning of trunks to handle Monday morning traffic while the rest of the week trunks are not fully utilized. Reducing the actual number of the trunks is contingent on moving the ACD peak demand off of Monday and then observing how traffic behaves on a typically busy Monday morning without the ACD traffic peaks.

A redialing model would be very useful. The survey of callers conducted for several Monday mornings at the PCARE site showed that most callers did not get through on their first attempts. Not only would a redialing model help to predict the average number of tries needed for success during heavy traffic, but it would allow a better estimate of demand. A previous study of traffic at the medical center showed traffic having a morning and afternoon peak, and when efforts were made to answer more calls in the morning, the afternoon peak faded away (34). Thus, when calls were answered the first time, callers did not need to retry later. This finding might also lend leverage for using answering machines to quickly inform callers when the

appointment computers are not operating or when all the appointments for a booking period are gone.

Issues related to caller perception require more of a survey approach or some sort of qualitative measure instead of using telephone traffic data. Is perception hurt worse by many busy signals or by reneging? If the caller never gets in (interflow), the worst that can be said is that calls can not get in, but if the call never gets answered (reneges), the caller has to wait and then can not get in. The calls answered criteria of this study only considered calls which were answered as getting into the system. Is an answering machine a valid option to manage perception? When a caller finally does get answered, the clerk tells the caller that access to an appointment is denied so the caller really is not in the system as they desired. The clerk would then inform the caller to call back next week to compete for an appointment which might open over three weeks from then. A machine could do this and do it quicker, but the effect on the perception of accessibility of the callers for this method for handling calls is unknown.

VII. CONCLUSIONS AND RECOMMENDATIONS

The initial problem for study has been decomposed into four parts: managing caller perception; setting ACD queue lengths; sizing the WHMC phone system for trunks; and, increasing the number of calls answered. However, the number of appointments available has risen as the hidden queue of the telephone access system. The above elements interact dynamically through the assumptions in the study which are based on the caller's desire for an appointment and the medical center's need to operate efficiently under expansion constraints and excess demand conditions. Alternatives for consideration are based on appointment clerk utilization and a measure of effectiveness which only gives significance to calls which get answered and to the reduction of blocking (overflow) in the system. The bottom line is that within the current system demand can be forecasted by ACD site, a management control structure over operators can be put in place using ACD data, more calls can be answered with shorter caller waiting times, and clerk productivity can be increased. Further, the number of commercial telephone lines required to be leased can easily be decreased. However, conscious decisions must be made, and made public knowledge for caller expectations to match,

on how to deal with calls which arrive for appointments after all appointments are dispersed for a booking period.

Recapitulation

At the busiest day of the month for WHMC's phone system, twelve appointment clerks staff ACD appointment phones which can account for up to forty percent of the traffic into the WHMC phone system. At the peak of traffic, thousands of calls get blocked (overflow) from the system, while thousands more get into the system, but are blocked (interflow) at their destinations, the ACD appointment sites. Only those ACD sites which book appointments during this busiest hour are severely affected by interflow, but all ACD sites have queue lengths which allow more calls to be accepted than the limited manning can answer before callers renege. ACD sites booking appointments on other days experience interflow in a similar fashion as those which book on Monday, but they do not have to contend with the high Monday traffic going to non-ACD phones. Consolidation of appointment centers or getting additional manpower to staff the phones at a site is not currently permitted due to technical and organizational constraints.

Appointment clerks provide the first impression callers receive of WHMC, and often feel stress when forced to turn away calls for appointments once all appointments for a

booking period have been given out. Many clerks still use notebooks to log and track appointments, manually collect provider and patient statistics, and remain on the ACD phones regardless of appointment availability or light traffic conditions. No feedback system or standard of service for appointment clerk productivity in handling ACD calls is uniformly in place, and clerk supervisors do not have uniform service policies concerning staffing, interruptions, or productivity. Callers are uninformed of the wait to expect or the probability of getting an appointment if successful in getting through to an appointment clerk via the ACD appointment phone system.

Findings

Research and simulation experimentation revealed several properties of the WHMC phone system. Phone service and system sizing at WHMC, with the constraints of a not-for-profit organization, cannot be successfully compared to industry examples of telephone service as with an airline, an insurance company, the telephone company, or a bank. Demand traffic for WHMC and the ACD appointment sites can be forecasted, and peaks for the ACD sites can be made to occur on any day of the week based on when appointments are available. Calls blocked (interflow) at the ACD sites is the best indicator of demand for these sites even though

redialing by callers inflates estimates of true demand. The non-ACD traffic is highest on Mondays and can affect access to ACD sites (and vice versa) resulting in overflow of the system. Queue length at the ACD sites (currently at six) allows more callers into the system than can be answered before reneging of calls occurs; these calls waste the time they hold the line and keep other calls from using the line. Cost savings from leasing fewer trunks cannot be realized until decisions for allowable blocking standards and weekly traffic balancing are made and their impact observed.

Appointment clerks are the first point of contact between the WHMC and the caller, but their task often goes unrecognized. Speed of service to callers is related to the number of servers, the kind of appointment system used, and whether interruptions can take a clerk away from answering the phone. Clerks have an upper limit to how many calls they can answer in an hour. The number of calls answered is the measure of effectiveness used under this study's definition of accessibility, and it is related to the speed of service, the intensity of traffic, and caller reneging. Caller perception of access was not linked with any service standard, but satisfying caller expectations was found to be important in the literature review.

Alternatives

Several alternatives are presented next for improving the number of calls answered and for reducing system overflow in order of increasing level of effort, cost, and change to the present system. It is a strategic issue to remember that no matter what system is chosen, getting to speak with an appointment clerk is not equivalent to getting an appointment. All alternatives (including no change) will give out all available appointments so long as there is excess demand. However, the total cost for the number of telephone lines leased can be decreased and the number of calls answered can be increased easily even within the constraints placed on this problem. There are many decisions which can be made, ranging from the way ACD sites use their clerks to automating the appointment systems where log books are currently used. Alternatives for decision makers to examine for increasing telephone access to WHMC are presented below. Most likely, many of the alternatives presented below would be chosen to become a scenario for comprehensive and continuing improvement of access to WHMC and its busiest appointment sites. A sample scenario using many of the alternatives discussed below is at the end of this section.

Shorten the ACD Queue Length. Adjust the queue length downward to minimize reneging across the ACD sites. During

peaks in demand at the ACD sites when most, or all, of the six waiting spaces are full, the waiting times for calls can be long and reneging occurs more frequently. By reducing the queue length to the point where even at the busy periods the amount of reneging becomes very small, more lines are made available to the non-ACD calls to reduce overflow, and only those ACD calls which could be served at the given service rate are made to wait. This action frees congested telephone resources from calls which cannot be completed and shortens the time on-hold for the new queue length. Though the number of busy signals ("all circuits busy" or interflow) may increase slightly, the number of calls answered at an ACD site stays the same. For example, if the queue length for the four ACD sites which currently book on Monday were cut from 6 to 3, 12 lines which previously might have been tied-up would now be added to the system; the clerks would still have plenty of calls to answer and only a few more actual ACD callers would be generating calls. Cost savings due to lessening the required number of leased lines might also be realized. This can be easily enacted by having the communications staff push a button or two on the ACD hardware.

ACD Data Automation. Automate the ACD data and create an on-line database of ACD information. Automation of the ACD data by capturing the output from the ACD software on the SL-1 switch will save the hand transcription effort

currently being done. In addition, with a little computer programming daily and weekly reports of telephone activity can be automatically generated. The benefits of this action are many: 1) the cumbersome process of getting a full set of data would disappear; 2) a standing source of information would be available to forecast demand at each ACD site from historical data to identify excess demand and to aid in scheduling operators; 3) a feedback loop for operator performance could be instigated with daily data, using hourly performance criteria for calls answered, which could be fed back to ACD sites automatically at the end of each day through the electronic mail software on the mainframe computer; and, 4) long-term tracking of ACD site service parameters and interflow patterns could be used to check the impact of management initiatives and to build a profile of service which might be useful in letting the public know what level of service to expect by ACD site. The cost of this action is very small; all that would be required is limited software development (which has already begun) and memory space either on part of the mainframe or on floppy disks for a PC.

Change Booking Days. Move all ACD booking off of Monday. The lessened traffic on Monday will decrease overflow to an as yet speculative amount, but cost savings from needing to lease fewer commercial lines are quite possible. Additionally, ACD site managers will have more

flexible use of their clerks during the typically busy Monday morning, while clerks will not be asked to do both administrative and phone answering tasks at the same time. Clerks will have more time to block-off appointment slots on the automated appointment system software (normally the next week's appointment slots are not ready until Friday afternoon and the clerks must block-off appointments on Monday morning before opening the lines). All of these benefits could be realized simply by changing the day appointments can be made, so that no ACD site books on Monday. Advance publicity on the change and the reasons for it should also be made.

Commercial Telephone Lines. If nothing is changed in the current system, the Monday morning system blocking will continue unless lines are added to the system. The simulation model suggested that approximately 140 lines would accommodate all the traffic at the busiest hour of the month for the 4 April 1988 system demand. However, this is tantamount to provisioning the entire medical center's telephone requirements on 1 out of 40 hours in the workweek. Even the current configuration of 90 lines is geared toward satisfying the demand of a 3 to 4 hour surge on Monday morning. If the ACD queue length is lowered and/or changes are made to move ACD booking days off of Monday, the weekly traffic into WHMC should be better balanced and even less than 90 lines would be needed to fully service WHMC without

any system overflow. These two simple changes could greatly improve telephone access to WHMC and even save about \$50 a month per line not leased.

ACD Site Management. Using the existing Operating Instruction (15), require clerks to be scheduled against the peaks in demand. The scheduling should dedicate the clerk solely to the task of answering the phone for the duration of the busy period. This will give the clerk clear responsibility for the task and make the clerks accountable for their own activity. Scheduling in this manner will also eliminate interruptions, so the clerks will have more time to answer the phone in the scheduled hours and the number of calls answered should increase. When the forecasted demand peaks wane, the site could return to the current mode of allowing interruptions in the phone answering task for flexibility in using the clerks. The cost is simple, site managers must be concerned enough with answering the calls to lose some flexibility over their clerks when there are peaks in call activity to that site.

Publicize Standards. In order to let the callers know what to expect, the database for the ACD sites could be used to show the worst and least waits for reaching an appointment clerk by time of the workweek. Callers needing to reach the appointment clerk but who are not seeking a new patient appointment could then avoid the bad periods. Additionally, each ACD site could report how often and when

it runs out of appointments for a typical booking period, so callers have an idea of their chances of getting an appointment provided they get through to the appointment site. This may be a way to establish reliability with the callers and show top management concern with the issue of telephone access, but there could be political arguments for not giving some of this information to the callers. In any case, it is unclear if anything short of quickly and courteously giving the caller the appointment for which they called will improve the perception of the accessibility to health care at WHMC.

Clerk Training. Train all clerks who make appointments by phone in the proper courtesy, patient sensitivity, appointment software (or log book), and ACD telephone hardware. Further, stress the importance of the clerks' role as the official greeter to WHMC. Caller comfort should increase and it should be easier to cross-train clerks for work at other ACD sites for either augmented manning or career rotation purposes. The cost of doing this is the time of the clerks and instructors, but much of this effort is currently on-going at WHMC.

Automate All ACD Sites. A standard computer appointment system should be used at all ACD sites. Service times for making appointments will be quicker, and the computer can calculate many of the routine patient and doctor statistics currently done by hand. The

standardization of appointment software and hardware should make it easier to cross-train clerks. More importantly, the number of clerks working at a particular ACD site could then be expanded, provided clerks could be medically trained to work at different ACD sites, by simply adding computer lines to the site (other hardware is transportable) and putting in additional ACD phone equipment.

Increase Manpower. Getting more clerks is next to impossible, but using the current ones imaginatively can multiply the number of clerks available at each ACD site. ACD sites should be compared to find which sites have comparable tasks in regards to the medical and organizational knowledge required to make appointments. Where there is comparable work, these ACD sites should be encouraged to book on different days and to cooperate by lending clerk manpower to each other at their respective peaks in ACD telephone traffic. ACD hardware and computer facilities will be needed at each site to support the additional help. The number of calls answered should increase with a better ability to handle the peaks in demand, and the liaison should foster the sharing of ideas on how to better deal with callers. This alternative allows for a possible expansion of the manpower pool without changing current lines of authority, but each site could lose control of its operator for a hour or two on each day

of augmentation for the reward of having extra help on its busiest day.

Answering Machines. A conscious decision must be made on how to deal with callers when no appointments are available in a booking period. The current policy uses the clerks as counselors who tell patients of their options. Using answering machines to address calls after all appointments are full while building a short waiting list from which to fill cancellations and "no shows", would create less stress for the clerks and allow the clerks to do other administrative duties. In addition, the callers would more quickly receive the information, rather than having to wait in line to finally be told to call back or to go to another provider of health care. Even if this scenario is unpalatable, a special answering machine to take effect on the occasions when the computer appointment system is inoperable would be useful to let callers know what is happening and to free the clerks time. The cost of the hardware and the unknown effect on caller perception stand in the way of this alternative, but the time congestion of the lines would definitely decrease since callers would get the nonavailability information faster (a 30 second message) than if they waited to be serviced by a clerk.

Hire Civilians as Operators. Putting civilian operators at the decentralized ACD sites would lend a continuity factor to the task, but would not likely add to

the number of calls answered, unless clerks would also help to answer calls during traffic peaks. Hiring civilian operators would free more of the clerks' time to do other tasks and allow clerks to get more career experience. This is basically paying more for the clerk time currently available.

Table IX. Value of Action Matrix
(in Order of Increasing Effort and Cost)

Alternative	Measure of Interest				
	Calls Answered	Over-flow	Inter-flow	Time Clerk Hold Util.	
Shorten ACD Queue	/	-	+	-	/
Automate ACD Data and Use Feedback	+	/	/	-	+
Change Booking Day	+	-	/	-	+
Schedule Clerk Time	+	/	/	-	/
Publicize Standards	/	/	/	/	/
Clerk Training	+	/	/	-	+
Automate Sites	+	-	/	-	+
Increase Manpower by Augmented Matrix	+	-	-	-	/
Answering Machines	+	-	-	-	+
Hire Civilians	/	/	/	/	/

Summary of Alternatives. The alternatives and their impact on access and traditional queuing measures are summarized in Table IX. The trends of choosing an alternative are revealed as a '+' if that measure would tend to increase, a '-' if that measure would tend to decrease, and a '/' if little or no change is expected in a measure for choosing a particular alternative. For example, the

alternative for automating the ACD data and using feedback to the ACD sites should make the clerks more accountable for their performance and more calls would be answered than if no attention were paid to the clerks. However, even if all the clerks each answered 5 additional calls an hour it is doubtful that this would significantly affect overflow or interflow at the busy hour of the month. It has been shown though that interflow for an entire day can be reduced (if calls are answered the first time they do not create a second afternoon peak). If the feedback encouraged the clerks to work harder to answer more calls, then the caller time on-hold would decrease and the clerk utilization (Util. in Table IX) would increase. The interpretation of the remainder of Table IX follows in like fashion.

A Sample Access Improvement Scenario. Consciously decide to give out appointments as fast as possible while alleviating Monday morning congestion. Use clerks for other administrative duties once appointments are dispersed by switching on an answering machine instead of having the clerks counsel callers. Move all ACD booking days off of Monday to balance traffic across the week. Ensure that ACD sites with similar tasks where clerks are capable of being cross-trained book appointments on different days of the week. Computerize all ACD appointment systems with standard software procedures for making appointments and for

collecting routine statistics. Place enough computer lines and ACD telephone lines at each site to accommodate augmented staffing levels of two, three, or more clerks. Track demand traffic through the ACD software by the hour registering interflow as well as calls accepted and calls answered for use in forecasting demand traffic by ACD site and for use as feedback to ACD site supervisors and clerks (as in PCARE). Staff to handle peaks in traffic using augmented manpower for the duration of the peak or until all appointments are given out. Identify excessive peaks in traffic and change appointment availability opportunities to smooth them out (as in GYN August 1988).

Once all appointments are full, to relieve clerk stress and use clerks for other duties, activate a recording to tell callers "all appointments are full" for a booking period. Include other recorded information deemed to be important to a caller who can not get an appointment that booking period (as in OPTHL). Make a short list of callers who arrive after all appointments are full, and use this list to fill "no shows" and cancellations typically experienced at the ACD site's clinics. Give this list priority when next making appointments via calling them on the following Monday if they were not offered a "no show" or cancellation slot. Reduce the queue size for the ACD sites to the level where renegeing is infrequent. Measure and validate clerk response times and use feedback to maintain

minimum service standards for ACD appointment clerks (there is an upper limit of about 30 to 35 calls answered per clerk per hour with computerized appointment system in heavy traffic). Prohibit interruptions of clerks while at their appointment making tasks; Monday morning has been freed to give site managers full use of clerks for other duties while the answering machines field calls. Publicize service levels, waiting times, and procedures that can reliably be expected.

Conclusion

The study of telephone access to WHMC and its busiest appointment centers has led to a better understanding of the technical and organizational issues surrounding telephone access in the constraints of the WHMC environment. Under the criteria of answering calls and reducing overflow, several possible actions were discussed to increase access through the telephone appointment system to WHMC. Alternatives to exploit the findings of this study were presented which require no change in the current lines of authority over manpower resources. Some of these actions have already been taken and found effective. Most notably, a change in appointment policy greatly reduced interflow at the GYN ACD site, and use of ACD data feedback and clerk scheduling improved the number of calls answered at the

PCARE ACD site. Training of clerks to insure better quality in politeness and in handling difficult callers has been under way since before this study began. The major claims for improving the number of ACD calls answered go unproven, since they cannot easily be tested in the real system, but it would be simple to reduce overflow, and possibly save money, by shortening the ACD queue lengths (until reneging is negligible) and moving all ACD booking off of Monday. A decision for continuous improvement of access with top management support is needed to ensure the gains promised by any change in the system.

Bibliography

1. Albin, Susan L. "Delays for Customers from Different Arrival Streams to a Queue," Management Science, 32: 329-340 (March 1986).
2. Bell, Colin E. "Optimal Operation of an M/M/2 Queue with Removable Servers," Operations Research, 28: 1189-1204 (Sep-Oct 1980).
3. Benes, V. E. "Growth, Complexity, and Performance of Telephone Connecting Networks," The Bell System Technical Journal, 62: 499-539 (February 1983).
4. Bezdek, P. V. and J. P. Collins. "Trunking Systems," The Bell System Technical Journal, 62: 2345-2364 (September 1983).
5. Boldy, Duncan P. and P. Ciaran O'Kane. "Health Operational Research -- A Selected Overview," European Journal of Operations Research, 10: 1-9 (1982).
6. Box, George E. P. and Norman R. Draper. Empirical Model-Building and Response Surfaces. New York: John Wiley & Sons, 1987.
7. Correll, John T. "What's Bugging the Troops: A Special Report," Air Force Magazine, 71: 66-67 (June 1988).
8. Crabill, Thomas B., and others. "A Classified Bibliography of Research on Optimal Design and Control of Queues," Operations Research, 25: 219-232 (March-April 1977).
9. Department of the Air Force. Criteria for Design and Construction of Air Force Health Facilities. AFR 88-50. Washington: HQ USAF, 16 May 1986.
10. Department of the Air Force. Medical Administration: Accessibility to Appointment Clerk. SGH OI 168-8. Lackland AFB, Texas: Wilford Hall Medical Center (ATC), 12 May 1988.
11. Descloux, A. "On Overflow Processes of Trunk Groups with Poisson Inputs and Exponential Service Times," The Bell System Technical Journal, ??: 383-397

(March 1963).

12. Dickens, Justin H. "Bus Terminal Telephone Information Study: An Operational Analysis of Patron Service Standards," Report by the Operations Standards Division, The Port of New York Authority, New York 11, New York, November 1962.
13. Duffy, F. P. and R. A. Mercer. "A Study of Network Performance and Customer Behavior During Direct-Distance-Dialing Call Attempts in the U. S. A.," The Bell System Technical Journal, 57: 1-33 (January 1978).
14. Elldin, Anders. "Dimensioning for the Dynamic Properties of Telephone Traffic," Ericsson Technics, 23: 315-344 (1967).
15. Everitt, D. E. and T. Downs. "The Output of the M/M/s Queue," Operations Research, 32: 796-808 (Jul-Aug 1984).
16. Falin, Gennadij I. "Multichannel Queuing Systems with Repeated Calls Under High Intensity of Repetition," Journal of Information Processing and Cybernetics, 23: 37-47 (1987).
17. Fortuin, Leonard. "Performance Indicators - Why, Where, and How?," European Journal of Operational Research, 34: 1-9 (1988).
18. Foster, F. G. and H. G. Perros. "On the Blocking Process in Queue Networks," European Journal of Operational Research, 5: 276-283 (1980).
19. Franks, R. L. and R. W. Rishel. "Overload Model of Telephone Network Operation," The Bell System Technical Journal, 52: 1589-1615 (November 1973).
20. Fredericks, A. A. "Congestion in Blocking Systems -- A Simple Approximation Technique," The Bell System Technical Journal, 59: 805-827 (Jul-Aug 1980).
21. Fredericks, A. A. and G. A. Reisner. "Approximations to Stochastic Service Systems, with an Application to a Retrial Model," The Bell System Technical Journal, 58: 557-576 (March 1979).
22. Glazer, Amihai and Refael Hassin. "?/M/1: On the Equilibrium Distribution of Customer Arrivals," European Journal of Operational Research, 13: 146-150 (1983).

23. Globerson, Shlomo. "Manpower Planning for a Telephone Service Department," Interfaces, 9: 105-111 (November 1979).
24. Grassmann, Winfield. "Transient Solutions in Markovian Queues: An Algorithm for Finding Them and Determining Their Waiting-time Distributions," European Journal of Operational Research, 1: 396-402 (1977).
25. Harel, Arie. "Sharp Bounds and Approximations for the Erlang Delay and Loss Formulas," Management Science, 34: 959-972 (August 1988).
26. Harris, Carl M. and others. "Modeling the IRS Telephone Taxpayer Information System," Operations Research, 35: 504-523 (July-August 1987).
27. Heffes, H. "Analysis of First-Come First-Served Queuing Systems with Peaked Inputs," The Bell System Technical Journal, 52: 1215-1228 (September 1973).
28. Heffes, H. and J. M. Holtzman. "Peakedness of Traffic Carried by a Finite Trunk Group With Renewal Input," The Bell System Technical Journal, 52: 1617-1642 (November 1973).
29. Hill, D. W. "The Effects of Traffic Load Variation on Measurement Accuracy," The Bell System Technical Journal, 56: 561-573 (April 1977).
30. Hill, D. W. and S. R. Neal. "Traffic Capacity of a Probability-Engineered Trunk Group," The Bell System Technical Journal, 55: 831-842 (September 1976).
31. Hillier, Frederick S. and Gerald J. Lieberman. Introduction to Operations Research (Fourth Edition). Oakland, California: Holden-Day, Inc, 1986.
32. Hoffman, Karla L. and Patsy B. Suanders. Evaluation of the IRS Telephone Information System: Simulation and Analysis. NBSIR 81-2198. Washington DC: National Bureau of Standards, January 1981.
33. Hoffman, Karla L. and Carl M. Harris. "Estimation of Caller Retrial Rate for a Telephone Information System," European Journal of Operational Research, 27: 207-214 (1986).
34. Jacobson, Colonel James M., Assistant Administrator Wilford Hall Medical Center. Personal Interview. Wilford Hall Medical Center, Lackland AFB TX, 8 July

1988.

35. Jagerman, D. L. "Methods in Traffic Calculation," Bell System Technical Journal, 63: 1283-1310 (September 1984).
36. Jarvis, J. F. "Approximating the Equilibrium Behavior of Multi-Server Loss Systems," Management Science, 31: 235-239 (February 1985).
37. Jolley, W. M. and R. J. Harris. "Analysis and Optimal Design of Queuing Systems with Post-call Activity," Australian Telecommunications Review, 15: 12-26 (1981).
38. Kashper, A., S. M. Rocklin, and C. R. Szilag. "Effects of Day-to-Day Load Variation on Trunk Group Blocking," The Bell System Technical Journal, 61: 123-135 (February 1982).
39. Keilson, J., J. Cozzolino, and H. Young. "A Service System with Unfilled Requests Repeated," Operations Research, 16: 1126-1137 (Nov-Dec 1968).
40. Kelton, W. David and Averill M. Law. "The Transient Behavior of the M/M/s Queue, with Implications for Steady-State Simulation," Operations Research, 33: 378-396 (Mar-Apr 1985).
41. Kolesar, Peter. "Stalking the Endangered CAT: A Queuing Analysis of Congestion at Automatic Teller Machines," Interfaces, 14: 16-26 (Nov-Dec 1984).
42. Larson, Richard C. "Perspectives on Queues: Social Justice and the Psychology of Queuing," Operations Research, 35: 895-905 (Nov-Dec 1987).
43. Law, Averill M. "Conservation Equations and Their Application to Queuing Simulations," Operations Research, 25: 279-288 (Mar-Apr 1977).
44. Linder, R. W. "The Development of Manpower and Facilities Planning Methods for Airline Telephone Reservation Offices," Operational Research Quarterly, 20: 3-22 (March 1969).
45. Liu, K. S. "Direct Distance Dialing: Call Completion and Customer Retrial Behavior," The Bell System Technical Journal, 59: 295-311 (March 1980).
46. McKeown, Patrick G. "An Application of Queuing Analysis to the New York State Child Abuse and Maltreatment Register Telephone Reporting System,"

Interfaces, 9: 20-25 (May 1979).

47. Messerli, E. J. "An Approximation for the Variance of the UPCO Offered Load Estimate," The Bell System Technical Journal, 57: 2575-2587 (September 1978).
48. Minh, Do Le. "Simulating GI/G/k Queues in Heavy Traffic," Management Science, 33: 1192-1200 (September 1987).
49. Mok, Stephen J. and J. George Shanthikumar "A Transient Queuing Model for Business Office with Standby Servers," European Journal of Operational Research, 28: 158-174 (1987).
50. Neal, S. R. and A. Kuczura. "A Theory of Traffic-Measurement Errors for Loss Systems With Renewal Input," The Bell System Technical Journal, 52: 967-988 (Jul-Aug 1973).
51. Oudheusden, Dirk L. and Wu Wen-Jeng. "Telephone Operator Scheduling with a Fixed Number of Operators," European Journal of Operational Research, 11: 55-59 (1982).
52. Patterson, Sgt. C. NCIOC GYN Clinic. Telephone Interview. Wilford Hall Medical Center, San Antonio TX, 18 October 1986.
53. Pritsker, A. Alan B. Introduction to Simulation and SLAM II (Third Edition). New York: A Halsted Press Book, John Wiley & Sons, 1986.
54. Rising, Edward J., Robert Baron, and Barry Averill. "A Systems Analysis of a University-Health-Service Outpatient Clinic," Operations Research, 21: 1030-1047 (Sep-Oct 1973).
55. Rothkopf, Micheal H. and Paul Rech. "Perspectives on Queues: Combining Queues is Not Always Beneficial," Operations Research, 35: 906-925 (Nov-Dec 1987).
56. Schruben, Lee W. and Barry H. Margolin. "Psuedorandom Number Assignment in Statistically Designed Simulation and Distribution Sampling Experiments," Journal of the American Statistical Association, 73: 504-520 (September 1976).
57. Sellers, Patricia. "How to Handle Customers' Gripes," Fortune, 18: 88-100 (October 1988).
58. Shannon, Robert E. "Simulation: A Survey with

Research Suggestions," AIIE Transactions: 289-301 (September 1975).

59. Shanthikumar, J. George, and David D. Yao. "On Server Allocation in Multiple Center Manufacturing Systems," Operations Research, 36: 333-341 (March-April 1988).
60. Sharma, R. L. and P. T. de Sousa. "Planning of Multi-Node Private Networks Using Automatic Call Distribution Systems," Proceedings of the IEEE, 44.5.1-5 (1981).
61. Smith, D. R. and W. Whitt. "Resource Sharing For Efficiency in Traffic Systems," The Bell System Technical Journal, 60: 39-55 (January 1981).
62. Szarkowicz, Donald S. and Thomas W. Knowles. "Optimal Control of an M/M/S Queuing System," Operations Research, 33: 644-660 (May Jun 1985).
63. Sze, David Y. "A Queuing Model for Telephone Operator Staffing," Operations Research, 32:229-248 (March-April 1984).
64. Weston, Frederick C. "Optimal Configuration of Telephone Answering Sites in a Service Industry," European Journal of Operational Research, 10: 395-405 (1982).
65. Whitaker, B. A. "Analysis and Optimal Design of a Multiserver, Multiqueue System with Finite Waiting Space in Each Queue," The Bell System Technical Journal, 54: 595-623 (March 1975).
66. Wilson, E. J. G. and R. J. Willis. "Scheduling of Telephone Betting Operators - A Case Study," Journal of the Operational Research Society, 34: 991-998 (1983).

Appendix A: A Sample ACD Hourly Report

Appendix B: SLAM Code with Fortran Subroutines

```
GEN,ELLS,THESISQ,7/27/1988,10//N//N//72;
LIMITS,12,4,500;
STAT,1,TIME IN S PCAR;
STAT,2,TIME IN S ORTH;
STAT,3,TIME IN S GYN;
STAT,4,TIME IN S IMED;
STAT,5,TIME IN S OPTH;
STAT,6,TIME IN S OB;
STAT,7,TIME IN S PEDS;
STAT,8,TIME IN S DERM;
STAT,9,TIME IN S SURG;
STAT,10,TIME IN S CARD;
STAT,11,TIME IN S NACD;
STAT,21,TIME HOLD PCAR;
STAT,22,TIME HOLD ORTH;
STAT,23,TIME HOLD GYN;
STAT,24,TIME HOLD IMED;
STAT,25,TIME HOLD OPTH;
STAT,26,TIME HOLD OB;
STAT,27,TIME HOLD PEDS;
STAT,28,TIME HOLD DERM;
STAT,29,TIME HOLD SURG;
STAT,30,TIME HOLD CARD;
TIMST,XX(1),NO. RENEGE PCAR;
TIMST,XX(2),NO. RENEGE ORTH;
TIMST,XX(3),NO. RENEGE GYN;
TIMST,XX(4),NO. RENEGE IMED;
TIMST,XX(5),NO. RENEGE OPTH;
TIMST,XX(6),NO. RENEGE OB;
TIMST,XX(7),NO. RENEGE PEDS;
TIMST,XX(8),NO. RENEGE DERM;
TIMST,XX(9),NO. RENEGE SURG;
TIMST,XX(10),NO. RENEGE CARD;
TIMST,XX(12),NO. FAST BUSY;
TIMST,XX(21),INTERFLOW PCAR;
TIMST,XX(22),INTERFLOW ORTH;
TIMST,XX(23),INTERFLOW GYN;
TIMST,XX(24),INTERFLOW IMED;
TIMST,XX(25),INTERFLOW OPTH;
TIMST,XX(26),INTERFLOW OB;
TIMST,XX(27),INTERFLOW PEDS;
TIMST,XX(28),INTERFLOW DERM;
TIMST,XX(29),INTERFLOW SURG;
TIMST,XX(30),INTERFLOW CARD;
TIMST,XX(31),NO. BUSY NACD;
TIMST,XX(41),CALLS ACPTD PCARE;
```

```

TIMST,XX(42),CALLS ACPTD ORTH;
TIMST,XX(43),CALLS ACPTD GYN;
TIMST,XX(44),CALLS ACPTD IMED;
TIMST,XX(45),CALLS ACPTD OPTH;
TIMST,XX(46),CALLS ACPTD OB;
TIMST,XX(47),CALLS ACPTD PEDS;
TIMST,XX(48),CALLS ACPTD DERM;
TIMST,XX(49),CALLS ACPTD SURG;
TIMST,XX(50),CALLS ACPTD CARD;
TIMST,XX(51),CALLS ACPTD NACD;
TIMST,XX(98),NACD CALL MINUTES;
TIMST,XX(99),CALL MINUTES;
; SEEDS STATEMENT
; INTLC SET THE RENEGE RATES FOR THE DIFFERENT SITES
; USE THE ACTUAL AVERAGE FOR THE TIME TO
; RENEGE SINCE NOTHING IS KNOWN ABOUT THE
; ACTUAL RENEGE DISTRIBUTION.
INTLC,XX(61)=0.863,XX(62)=1.783,XX(63)=3.4167,XX(64)=0.5667;
INTLC,XX(65)=0.610,XX(66)=2.200,XX(67)=2.5833,XX(68)=2.1600;
INTLC,XX(69)=1.1367,XX(70)=1.3833,XX(71)=15.0;
; ADD LOGIC FOR VARYING HOURLY RATES
NETWORK;
    RESOURCE/12,TRUNK(90),12;
    RESOURCE/1,PCARE(2),1;
    RESOURCE/2,ORTH(1),2;
    RESOURCE/3,GYN(1),3;
    RESOURCE/4,INMED(1),4;
    RESOURCE/5,OPTH(1),5;
    RESOURCE/6,OBST(1),6;
    RESOURCE/7,PEDS(2),7;
    RESOURCE/8,DERM(1),8;
    RESOURCE/9,GSURG(1),9;
    RESOURCE/10,CARD(1),10;
    RESOURCE/11,NONACD(100),11;
;
; CREATE ARRIVALS TO THE CITY TRUNK SYSTEM TO COMPETE
; FOR ENTRY TO WHMC. SET THE ARRIVAL RATE AND THE TIME
; THE CLINIC OPENS FOR BUSINESS. MARK ATTRIBUTE 1 AS
; TIME ENTERED SYSTEM. USE DIFFERENT SEEDS IN RV
; GENERATION.
; MARK ATTRIBUTE 2 AS THE FILE NUMBER FOR THE RESPECTIVE
; CLINIC OF DESTINATION.
;
; PRIMARY CARE
;
    CREATE,EXPON(0.4140,1),0,1;
    ASSIGN,ATRIB(2)=1;
    ACT,,,CITY;
    TERM;
; ORTHOPEDICS
    CREATE,EXPON(0.1500,2),0,1;
    ASSIGN,ATRIB(2)=2;

```

```

        ACT,,,CITY;
        TERM;
;   GYN
        CREATE,EXPON(0.0360,3),0,1;
        ASSIGN,ATRIB(2)=3;
        ACT,,,CITY;
        TERM;
;   INTERNAL MEDICINE
        CREATE,EXPON(1.9260,1),0,1;
        ASSIGN,ATRIB(2)=4;
        ACT,,,CITY;
        TERM;
;   OPHTHALMOLOGY
        CREATE,EXPON(1.7460,2),0,1;
        ASSIGN,ATRIB(2)=5;
        ACT,,,CITY;
        TERM;
;   OB
        CREATE,EXPON(2.4180,3),0,1;
        ASSIGN,ATRIB(2)=6;
        ACT,,,CITY;
        TERM;
;   PEDIATRICS
        CREATE,EXPON(0.8700,1),0,1;
        ASSIGN,ATRIB(2)=7;
        ACT,,,CITY;
        TERM;
;   DERMATOLOGY
        CREATE,EXPON(1.3620,2),0,1;
        ASSIGN,ATRIB(2)=8;
        ACT,,,CITY;
        TERM;
;   GENERAL SURGERY
        CREATE,EXPON(1.7520,3),0,1;
        ASSIGN,ATRIB(2)=9;
        ACT,,,CITY;
        TERM;
;   CARDIOLOGY
        CREATE,EXPON(2.7000,1),0,1;
        ASSIGN,ATRIB(2)=10;
        ACT,,,CITY;
        TERM;
;   NON ACD
        CREATE,EXPON(0.01156,2),0,1;
        ASSIGN,ATRIB(2)=11;
        ACT,,,CITY;
        TERM;
;
;   ARRIVAL TO CITY TRUNK TO COMPETE FOR LINE TO WHMC
;   IF NOT CONNECT, THEN GET FAST BUSY (FBZY) SIGNAL
;   IF CONNECT, SEND TO EVENT NODE TO SET RENEGE TIME
;   THEN SEND INSTANTANEOUSLY TO DESTINATION CLINIC

```

```

;
CITY AWAIT(12/1),TRUNK/1,BALK(FBZY),1;
    EVENT,1,1;
    TERM;

;
;   ADD OTHER ROUTES TO OTHER CLINICS
;   INCREMENT COUNTER XX(12) FOR NUMBER OF BLOCKS
;   THEN THROW ENTITY OUT OF SYSTEM
;   COLLECT ON TIME BETWEEN FAST BUSY (BLOCKED) ATTEMPTS
;
FBZY COLCT,BET,TIME BET. FBZY;
    ASSIGN,XX(12)=XX(12)+1;
    TERM;

;
;   ARRIVE AT PRIMARY CARE CLINIC ACD SITE TO TRY TO JOIN
QUEUE
;   IF QUEUE IS FULL THEN GET ALL CIRCUITS BUSY MSG (SBZ#)
;   IF GET SERVER, DROP RENEGE EVENT PREVIOUSLY SCHEDULED
;   COLLECT INFO ON TIME IN SYSTEM AND TIME BETWEEN
DEPARTURES
;
PCAR AWAIT(1/6),PCARE/1,BALK(SBZ1),1;
    EVENT,3;
    COLCT(21),INT(1),TIME HOLD PCAR;
    ACTIVITY/1,EXPON(1.36,4);
    COLCT(1),INT(1),TIME IN S PCAR;
    FREE,TRUNK/1,1;
    ASSIGN,XX(99)=XX(99)+TNOW-ATRI(1);
    ACT/21,EXPON(0.3667,4);
    FREE,PCARE/1,1;
    TERM;

;
;   OVERFLOW AT THE QUEUE TO ALL CIRCUITS BUZY MESSAGE
;   INCREMENT COUNTER XX(21) FOR OVERFLOW Q1
;
SBZ1 EVENT,3;
    FREE,TRUNK/1,1;
    ASSIGN,XX(21)=XX(21)+1;
    TERM;

;
;   ARRIVE AT ORTHOPEDICS CLINIC ACD SITE
;
ORTH AWAIT(2/6),ORTH/1,BALK(SBZ2),1;
    EVENT,3;
    COLCT(22),INT(1),TIME HOLD ORTH;
    ACT/2,EXPON(2.4333,5);
    COLCT(2),INT(1),TIME IN S ORTH;
    FREE,TRUNK/1,1;
    ASSIGN,XX(99)=XX(99)+TNOW-ATRI(1);
    ACT/22,EXPON(0.2933,5);
    FREE,ORTH/1,1;
    TERM;

```

```

;
;   ALL CIRCUITS BUSY MESSAGE AT ORTHOPEDICS
;
SBZ2 EVENT,3;
    FREE,TRUNK/1,1;
    ASSIGN,XX(22)=XX(22)+1;
    TERM;
;
;   ARRIVE AT GYNECOLOGY CLINIC ACD SITE
;
GY   AWAIT(3/6),GYN/1,BALK(SBZ3),1;
    EVENT,3;
    COLCT(23),INT(1),TIME HOLD GYN;
    ACT/3,EXPON(2.000,6);
    COLCT(3),INT(1),TIME IN S GYN;
    FREE,TRUNK/1,1;
    ASSIGN,XX(99)=XX(99)+TNOW-ATIB(1);
    ACT/23,EXPON(0.4233,6);
    FREE,GYN/1,1;
    TERM;
;
;   ALL CIRCUITS BUSY MESSAGE AT GYN
;
SBZ3 EVENT,3;
    FREE,TRUNK/1,1;
    ASSIGN,XX(23)=XX(23)+1;
    TERM;
;
;   ARRIVE AT INTERNAL MEDICINE CLINIC
;
IMED AWAIT(4/6),INMED/1,BALK(SBZ4),1;
    EVENT,3;
    COLCT(24),INT(1),TIME HOLD IMED;
    ACTIVITY/4,EXPON(1.7967,4);
    COLCT(4),INT(1),TIME IN S IMED;
    FREE,TRUNK/1;
    ASSIGN,XX(99)=XX(99)+TNOW-ATIB(1);
    ACT/24,EXPON(0.2200,4);
    FREE,INMED/1;
    TERM;
;
;   ALL CIRCUITS BUSY MESSAGE INTERNAL MEDICINE
;
SBZ4 EVENT,3;
    FREE,TRUNK/1;
    ASSIGN,XX(24)=XX(24)+1;
    TERM;
;
;   ARRIVE AT OPHTHALMOLOGY
;
OPHTH AWAIT(5/6),OPHTH/1,BALK(SBZ5),1;
    EVENT,3;

```

```

COLCT(25),INT(1),TIME HOLD OPTH;
ACTIVITY/5,EXPON(1.1833,5);
COLCT(5),INT(1),TIME IN S OPTH;
FREE,TRUNK/1;
ASSIGN,XX(99)=XX(99)+TNOW-ATRIB(1);
ACT/25,EXPON(0.0333,5);
FREE,OPTHL/1;
TERM;

;
;   ALL CIRCUITS BUSY MESSAGE FROM OPHTHALMOLOGY
;
SBZ5 EVENT,3;
FREE,TRUNK/1;
ASSIGN,XX(25)=XX(25)+1;
TERM;

;
;   ARRIVE OBSTETRICS
;
OB   AWAIT(6/6),OBST/1,BALK(SBZ6),1;
EVENT,3;
COLCT(26),INT(1),TIME HOLD OB;
ACTIVITY/6,EXPON(1.2117,6);
COLCT(6),INT(1),TIME IN S OB;
FREE,TRUNK/1;
ASSIGN,XX(99)=XX(99)+TNOW-ATRIB(1);
ACT/26,EXPON(2.8833,6);
FREE,OBST/1;
TERM;

;
;   ALL CIRCUITS BUSY MESSAGE OBSTETRICS
;
SBZ6 EVENT,3;
FREE,TRUNK/1;
ASSIGN,XX(26)=XX(26)+1;
TERM;

;
;   ARRIVE PEDIATRICS
;
PDS   AWAIT(7/6),PEDS/1,BALK(SBZ7),1;
EVENT,3;
COLCT(27),INT(1),TIME HOLD PEDS;
ACTIVITY/7,EXPON(1.8667,4);
COLCT(7),INT(1),TIME IN S PEDS;
FREE,TRUNK/1;
ASSIGN,XX(99)=XX(99)+TNOW-ATRIB(1);
ACT/27,EXPON(1.7667,4);
FREE,PEDS/1;
TERM;

;
;   ALL CIRCUITS BUSY MESSAGE PEDIATRICS
;
SBZ7 EVENT,3;

```

```

    FREE,TRUNK/1;
    ASSIGN,XX(27)=XX(27)+1;
    TERM;
;
;   ARRIVE DERMATOLOGY
;
DRM  AWAIT(8/6),DERM/1,BALK(SBZ8),1;
      EVENT,3;
      COLCT(28),INT(1),TIME HOLD DERM;
      ACTIVITY/8,EXPON(2.0667,5);
      COLCT(8),INT(1),TIME IN S DERM;
      FREE,TRUNK/1;
      ASSIGN,XX(99)=XX(99)+TNOW-ATRIB(1);
      ACT/28,EXPON(0.4667,5);
      FREE,DERM/1;
      TERM;
;
;   ALL CIRCUITS BUSY DERMATOLOGY
;
SBZ8 EVENT,3;
      FREE,TRUNK/1;
      ASSIGN,XX(28)=XX(28)+1;
      TERM;
;
;   ARRIVE GENERAL SURGERY
;
GSRG AWAIT(9/6),GSURG/1,BALK(SBZ9),1;
      EVENT,3;
      COLCT(29),INT(1),TIME HOLD SURG;
      ACTIVITY/9,EXPON(0.8667,6);
      COLCT(9),INT(1),TIME IN S SURG;
      FREE,TRUNK/1;
      ASSIGN,XX(99)=XX(99)+TNOW-ATRIB(1);
      ACT/29,EXPON(0.5000,6);
      FREE,GSURG/1;
      TERM;
;
;   ALL CIRCUITS BUSY GENERAL SURGERY
;
SBZ9 EVENT,3;
      FREE,TRUNK/1;
      ASSIGN,XX(29)=XX(29)+1;
      TERM;
;
;   ARRIVE CARDIOLOGY
;
CRD  AWAIT(10/6),CARD/1,BALK(SBZ0),1;
      EVENT,3;
      COLCT(30),INT(1),TIME HOLD CARD;
      ACTIVITY/10,EXPON(2.2167,4);
      COLCT(10),INT(1),TIME IN S CARD;
      FREE,TRUNK/1;

```

[illegible]


```
MONTR,CLEAR,5;
FIN;
```

```
PROGRAM MAIN
DIMENSION NSET(10000)
INCLUDE 'SLAM$DIR:PARAM.INC'
COMMON/SCOM1/ATRIB(MATRB), DD(MEQT), DDL(MEQT),
DTNOW, II, MFA, 1MSTOP,NCLNR, NCRDR, NPRNT,
NNRUN, NNSET, NTAPE, SS(MEQT),
2SSL(MEQT),TNEXT, TNOW, XX(MMXXV)
COMMON QSET(10000)
EQUIVALENCE (NSET(1),QSET(1))
OPEN (UNIT=8,FILE='THEWHMC.DAT',STATUS='NEW')
NNSET=10000
NCRDR=5
NPRNT=6
NTAPE=7
NFLOT=2
CALL SLAM
STOP
```

C

```
END
```

```
SUBROUTINE EVENT(I)
INCLUDE 'SLAM$DIR:PARAM.INC'
COMMON/SCOM1/ATRIB(MATRB), DD(MEQT), DDL(MEQT),
DTNOW, II, MFA, 1MSTOP,NCLNR, NCRDR, NPRNT,
NNRUN, NNSET, NTAPE, SS(MEQT),
2SSL(MEQT),TNEXT, TNOW, XX(MMXXV)
GO TO (1,2,3),I
```

C

```
C *** CODE FOR RENEGING FROM A QUEUE USING NETWORK CALLS
C *** DEFINE EVENT CODE 1 AS SCHEDULING A POTENTIAL RENEGE
C ***
C *** MPREN IS A POINTER TO THE RENEGE EVENT
C *** MPCQ IS A POINTER TO THE CALLER IN THE QUEUE
```

C

```
1 MPREN=MFA
MPCQ=NSUCR(MFA)
```

C

```
C *** SET ATRIB(3) OF CALLER WITH POTENTIAL RENEGE EVENT
C *** TO THE POINTER FOR THE RENEGE EVENT
```

C

```
ATRIB(3)=REAL(MPCQ)
```

C

```
C *** SCHEDULE A FUTURE CALL OF THE RENEGE EVENT AND
C *** REFERENCE EVENT CODE 2 (THE ACTUAL RENEGE: BELOW)
```

C

```
IF(ATRIB(2).EQ.1) CALL SCHDL(2,EXPON(XX(61),3),ATRIB)
IF(ATRIB(2).EQ.2) CALL SCHDL(2,EXPON(XX(62),3),ATRIB)
```

```

      IF(ATTRIB(2).EQ.3) CALL SCHDL(2,EXPON(XX(63),3),ATTRIB)
      IF(ATTRIB(2).EQ.4) CALL SCHDL(2,EXPON(XX(64),3),ATTRIB)
      IF(ATTRIB(2).EQ.5) CALL SCHDL(2,EXPON(XX(65),3),ATTRIB)
      IF(ATTRIB(2).EQ.6) CALL SCHDL(2,EXPON(XX(66),3),ATTRIB)
      IF(ATTRIB(2).EQ.7) CALL SCHDL(2,EXPON(XX(67),3),ATTRIB)
      IF(ATTRIB(2).EQ.8) CALL SCHDL(2,EXPON(XX(68),3),ATTRIB)
      IF(ATTRIB(2).EQ.9) CALL SCHDL(2,EXPON(XX(69),3),ATTRIB)
      IF(ATTRIB(2).EQ.10) CALL SCHDL(2,EXPON(XX(70),3),ATTRIB)
      IF(ATTRIB(2).EQ.11) CALL SCHDL(2,EXPON(XX(71),3),ATTRIB)
C
C *** SET ATTRIB(4) OF THE CALLER TO THE POINTER FOR THE
C *** RENEGE EVENT AND PUT THEM IN THE FILE FOR THE ACD SITE
C
      ATTRIB(4)=REAL(MPREN)
      INQUE=INT(ATTRIB(2))
      CALL FILEM(INQUE,ATTRIB)
      ITEMP=INQUE+40
      XX(ITEMP)=XX(ITEMP)+1
      RETURN
C
C *** DEFINE EVENT 2 AS THE ACTUAL RENEGE EVENT
C ***
C *** REMOVE THE CALLER FROM WHATEVER AWAIT NODE FILE
C *** THEY ARE IN (ATTRIB(2)) A THE COUNTER FOR
C *** RENEGING FROM THAT QUEUES INCREMENTED
C ***
C *** WCHQ IS "WHICH QUEUE" AND IS SET BY ATTRIB(2)
C *** MPCQ IS SET TO THE POINTER OF THE RENEGE EVENT
C *** (ATTRIB(3))
C
      2  IWCHQ=INT(ATTRIB(2))
      MPCQ=INT(ATTRIB(3))
      CALL RMOVE(-MPCQ,IWCHQ,ATTRIB)
      XX(IWCHQ)=XX(IWCHQ)+1
      CALL FREE(12,1)
      RETURN
C
C *** DEFINE EVENT 3 AS CANCELING A RENEGE EVENT
C *** WHEN THE CALLER HAS FINALLY REACHED A SERVER
C *** SUCCESSFULLY
C *** TAKE THE FUTURE RENEGE EVENT OFF THE CALENDAR
C
      3  MPREN=INT(ATTRIB(4))
      CALL RMOVE(-MPREN,NCLNR,ATTRIB)
      RETURN
      END
C
C
      SUBROUTINE OPUT
      INCLUDE 'SLAM$DIR:PARAM.INC'
      COMMON/SCOM1/ATTRIB(MATRB), DD(MEQT), DDL(MEQT),
      DTNOW, II, MFA, 1MSTOP,NCLNR, NCRDR, NPRNT,

```

```

      NNRUN, NNSET, NTAPE, SS(MEQT),
      2SSL(MEQT),TNEXT, TNOW, XX(MMXXV)
C
C *** PRINT REPORT TO OUTPUT FILE 'THEWHMC.DAT'
C
      WRITE(8,10)'  NUMBER OF CITY BLOCKINGS = ',XX(12)
10  FORMAT(A27,F8.2)
      SUMREN=0.0
      DO 20 I=1,10
          SUMREN=SUMREN+XX(I)
20  CONTINUE
      WRITE(8,25)'  TOTAL NUMBER OF RENEGES = ',SUMREN
25  FORMAT(A36,F7.0)
      SUMSBZ=0.0
      DO 30 J=21,30
          SUMSBZ=SUMSBZ+XX(J)
30  CONTINUE
      WRITE(8,35)'  TOTAL INTERFLOW (BUSY MESSAGE) =
          ',SUMSBZ
35  FORMAT(A51,F7.2)
      ACCEPT=0.0
      DO 40 K=41,51
          ACCEPT=ACCEPT+XX(K)
40  CONTINUE
      WRITE(8,45)'  TOTAL CALLS ACCEPTED = ',ACCEPT
45  FORMAT(A33,F8.2)
      WRITE(8,50)'  CALL MINUTES FOR NACD AND ACD
          ARE',XX(98),XX(99)
50  FORMAT(A40,2F8.2)
      WRITE(8,55)'
55  FORMAT(A40)
      RETURN
      END

```

VITA

Captain Martin E. Ellingsworth was born [REDACTED]

[REDACTED] He graduated from high school in [REDACTED] and attended the United States Air Force Academy, from which he received the degree of Bachelor of Science in Operations Research in May 1984. Upon graduation, he received a regular commission in the USAF. He served as a Personnel Operations Research Analyst with the Air Force Human Resources Laboratory, Brooks AFB, Texas until going to Squadron Officer School en route to entering the School of Engineering, Air Force Institute of Technology, in June of 1987.

[REDACTED]

[REDACTED]

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE

AD+202709

REPORT DOCUMENTATION PAGE				Form Approved OMB No. 0704-0188	
1a. REPORT SECURITY CLASSIFICATION UNCLASSIFIED			1b. RESTRICTIVE MARKINGS		
2a. SECURITY CLASSIFICATION AUTHORITY			3. DISTRIBUTION/AVAILABILITY OF REPORT Approved for public release; distribution unlimited.		
2b. DECLASSIFICATION/DOWNGRADING SCHEDULE					
4. PERFORMING ORGANIZATION REPORT NUMBER(S) AFIT/GOR/ENS/88D-6			5. MONITORING ORGANIZATION REPORT NUMBER(S)		
6a. NAME OF PERFORMING ORGANIZATION School of Engineering		6b. OFFICE SYMBOL (If applicable) AFIT/ENS	7a. NAME OF MONITORING ORGANIZATION		
6c. ADDRESS (City, State, and ZIP Code) Air Force Institute of Technology Wright-Patterson AFB OH 45433-6583			7b. ADDRESS (City, State, and ZIP Code)		
8a. NAME OF FUNDING/SPONSORING ORGANIZATION		8b. OFFICE SYMBOL (If applicable)	9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER		
8c. ADDRESS (City, State, and ZIP Code)			10. SOURCE OF FUNDING NUMBERS		
			PROGRAM ELEMENT NO.	PROJECT NO.	TASK NO.
11. TITLE (Include Security Classification) See Box 19					
12. PERSONAL AUTHOR(S) Martin E. Ellingsworth, B.S., Capt., USAF					
13a. TYPE OF REPORT MS Thesis		13b. TIME COVERED FROM _____ TO _____		14. DATE OF REPORT (Year, Month, Day) 1988 December	
				15. PAGE COUNT 172	
16. SUPPLEMENTARY NOTATION					
17. COSATI CODES			18. SUBJECT TERMS (Continue on reverse if necessary and identify by block number) Telephone Systems, Queuing Theory, Healthcare Facilities, Digital Simulation		
FIELD	GROUP	SUB-GROUP			
25	03				
06	12				
19. ABSTRACT (Continue on reverse if necessary and identify by block number) Title: MODELING TELEPHONE ACCESS TO WILFORD HALL MEDICAL CENTER AND ITS BUSIEST APPOINTMENT SITES Thesis Chairman: Joseph R. Litko, Major, USAF Assistant Professor Department of Operational Sciences <i>88 Arsenic media 12 Jan 1999</i>					
20. DISTRIBUTION/AVAILABILITY OF ABSTRACT <input type="checkbox"/> UNCLASSIFIED/UNLIMITED <input checked="" type="checkbox"/> SAME AS RPT. <input type="checkbox"/> DTIC USERS			21. ABSTRACT SECURITY CLASSIFICATION UNCLASSIFIED		
22a. NAME OF RESPONSIBLE INDIVIDUAL Joseph R. Litko, Major, USAF			22b. TELEPHONE (Include Area Code) (513) 255-3362		22c. OFFICE SYMBOL AFIT/ENS

DD Form 1473, JUN 86

Previous editions are obsolete.

SECURITY CLASSIFICATION OF THIS PAGE

UNCLASSIFIED

19. Telephone traffic into Wilford Hall Medical Center and its busiest appointment centers was simulated at the busiest hour of the month to explore alternatives for improving telephone access to health care. Telephone access to care was defined in two parts: 1) decreasing calls blocked (overflow) of the central commercial switch; and, 2) increasing the number of calls answered at the decentralized appointment sites, which operate with automatic call distributors. Methods for improving current call data reporting procedures were developed. Avenues for feeding back operator performance data to the appointment sites were constructed which may serve as a management control system for monitoring continuous and sustained improvement in the telephone access at each appointment site. The simulation gave insight into how to manage call traffic to decrease the number of commercial lines leased while ensuring a minimum of overflow traffic. Simple changes to the length of the queue of the call distributors and to when new appointments are made available could dramatically decrease the load on the network at the busiest hour of the month. Alternatives for increasing the number of calls answered included dedicated scheduling of operators, automating the appointment making process, and augmented staffing of operators between the decentralized sites. The strategic problem behind telephone access is that getting to talk to an appointment clerk is not equivalent to getting the appointment for which the call originated. A key question for management is how to deal with calls after all available appointments are distributed for a given clinic. Findings for managing demand traffic and operator performance to improve telephone access to care within the constraints of a not-for-profit health care organization may be useful to other such organizations.

Keywords: Health care
facilities, Digital
simulation,
Queueing theory,
Theses. (SEW/KT)